



seit 1558

Friedrich – Schiller – Universität Jena
Biologisch – Pharmazeutische Fakultät

**"Molecular and functional analysis of Or22a transcripts across
different *Drosophila ananassae* populations."**

Diplomarbeit
Zur Erlangung des akademischen Grades eines
Diplom – Biologen

accomplished at the:



Max Planck Institute
for Chemical Ecology

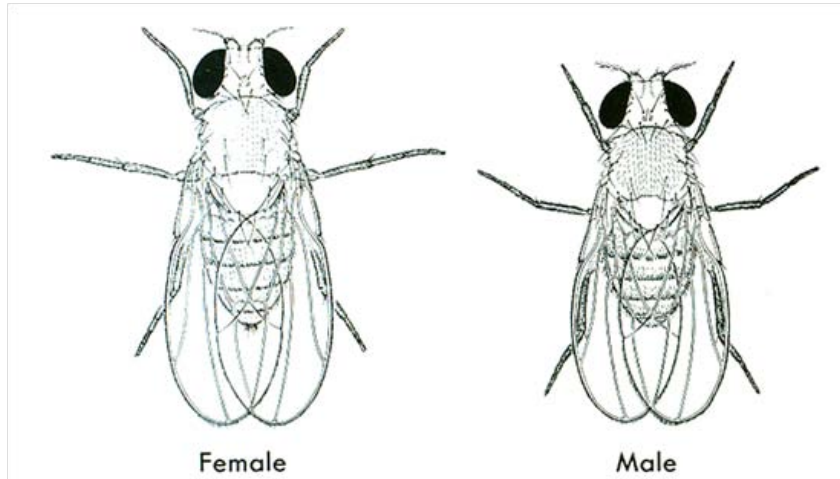


MAX-PLANCK-GESellschaft

submitted by

Anna Kretschmer

[November / 2011]



Drosophila ananassae

(Moriwaki 1936)

First Reviewer: Dr. Marcus Carl Stensmyr

Second Reviewer: Prof. Dr. Bill Hansson

Table of contents:

1. Introduction	1
1.1 The species <i>Drosophila ananassae</i>	1
1.2 Insect's sense of smell	4
1.3 Odorant receptors	6
1.3.1 Expression pattern of OR genes	7
1.3.2 The odorant receptor 22a	9
1.4 Purpose of the Study	10
2. Material	11
2.1 Fly stocks	11
2.2 Bacterial strain	11
2.3 Cloning vectors	11
2.4 Synthesized Oligonucleotides	11
2.5 Enzymes	13
2.6 Kits	13
2.7 DNA ladder/ size marker	14
2.8 Chemicals and consumables	14
2.9 Antibiotics	15
2.10 Buffers and solutions	15
2.11 Immunostaining	16
2.12 Laboratory equipment	17

3. Methods	19
3.1 DNA extraction	19
3.2 Primer design	19
3.3 PCR procedure	20
3.3.1 General mode of action	20
3.3.2 Master Mix and Polymerases	20
3.3.3 PCR conditions	21
3.4 Agarose gel electrophoresis	21
3.4.1 General mode of action	21
3.4.2 Separation and purification of PCR products	22
3.5 Transformation and selection	22
3.6 Colony PCR	25
3.7 Plasmid isolation	25
3.8 Sequencing and sequence analyses	26
3.8.1 Sequencing	26
3.8.2 Sequence analyses	26
3.8.3 Open reading frame (ORF) and translation into amino acid sequence	27
3.9 Phylogenetic analyses	27
3.10 Transmembrane domain prediction	28
3.11 Construct design for making a transgenic fly	28
3.12 Transgenic fly lines	31

3.13 Immunostaining	32
4. Results & Discussion	33
4.1 DNA Extraction	33
4.2 PCR amplification	34
4.3 Sequence analyses and alignment	35
4.3.1 <i>DanaOr22a-2</i>	35
4.3.2 <i>DanaOr22a-3</i>	37
4.3.3 <i>DanaOr22a-6</i>	37
4.4 Phylogenetic trees	37
4.4.1 Bayesian tree for the <i>Or</i> gene family	37
4.4.2 Phylogenetic tree based on maximum parsimony of <i>Or22a</i> genes of <i>D. ananassae</i>	38
4.5 Transmembrane domain predictions	40
4.5.1 <i>DanaOr22a-2</i>	40
4.5.2 <i>DanaOr22a-3</i> and <i>DanaOr22a-6</i>	42
4.5.3 Infinite island model	42
4.6 Immunostaining	44
5. Summary	46
6. Outlook	47
7. References	48
8. Declaration of academic honesty	52
9. Abbreviations	53
10. Acknowledgements	54

1. Introduction

1.1 The species *Drosophila ananassae*

Drosophila ananassae constitutes a sister group of all members of the melanogaster subgroup (*D. melanogaster*, *D. simulans*, *D. yakuba*, *D. sechellia*, *D. erecta*) within the melanogaster taxa (Clark *et al.* 2007) (see **Figure 1**). Like *D. melanogaster* and *D. simulans* *D. ananassae* is a cosmopolitan species but in contrast to these the population structure from *D. ananassae* is geographically highly subdivided. Geographical barriers like oceans or mountains provoked high-level structured populations, which qualified the species as a model system for population and evolutionary genetics (Das *et al.* 2004; Matsuda *et al.* 2009; Vogl *et al.* 2003). Especially such a highly structured population plays an important role for understanding how evolutionary forces work e.g., mutation, migration and drift (Baines *et al.* 2004; Vogl *et al.* 2003).

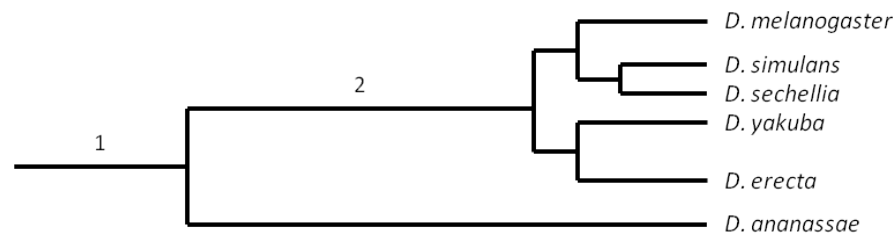


Figure 1 | Cladogram of 6 sequenced *Drosophila* species. Phylogenetic tree shows predicted relationships between members of the *melanogaster* group (1) and also between species of *melanogaster* subgroup (2). Figure adapted from (Clark *et al.* 2007).

The species *Drosophila ananassae* was first described in 1858 by Carl Ludwig Doleschall during his research in Ambon, Indonesia (Doleschall 1858). Thereafter, *D. ananassae* was recorded in tropical, subtropical and mildly tempered biogeographic zones all over the world (e.g. in Southeast Asia, West Africa, South America, U.S.A. in southern Texas and Florida, Northeast Australia and on many pacific islands). In colder temperate areas as in e.g. large parts of North America this species is absent (Bock & Wheeler 1972; J.T. & W.S. 1952; Tobari 1993) (see **Figure 2**). Thus, it is a cosmopolitan species with an ancestral species range as well as many genetically independent

peripheral populations (Das *et al.* 2004). Furthermore, studies of chromosomal variability (Dobzhansky & Dreyfus 1943) and recent analysis of DNA sequence polymorphisms (Das *et al.* 2004; Vogl *et al.* 2003) and microsatellite variation (Schug *et al.* 2004; Schug *et al.* 2007) have indicated Southeast Asia as the place of origin of *D. ananassae*. Outside of Southeast Asia the species is mostly found in connection to humans and appears rarely in natural habitats (Bock & Parsons 1978). Its ancestral species range is delimited to a region called Sundaland, which was a single landmass during the late Pleistocene (around 18,000 years ago). During that time period the sea level was around 120 m lower than today. Peripheral *D. ananassae* populations (e.g. those inhabiting parts of America, the islands of the Pacific Ocean and Australia) seem to have been spread by man, from their native point of origin and introduced into a new environment (Baines *et al.* 2004; Das *et al.* 2004; Dobzhansky & Dreyfus 1943; Vogl *et al.* 2003). Like all “domestic” species, *D. ananassae* conquered and filled ecological niches and territories that were made available by human activity (Dobzhansky 1965).

Infinite island model

Population subdivision was modelled several times on analytical and computational fronts. Relatively simple models like the island model (Maruyama 1970; Wright 1931) and the one – and two – dimensional stepping stone model (Kimura & Weiss 1964) have shed light on general processes of the complex demography of natural and domestic *D. ananassae* populations (Vogl *et al.* 2003; Wakeley 2001). The original island model assumes a “single parameter describing the correlation of two random gametes within a population” (Vogl *et al.* 2003) and a parameter called migration factor γ which has the same value for all populations in the estimated randomly mating species (Vogl *et al.* 2003; Wright 1931, 1969). Indeed, all natural and domestic populations differ mostly in both size and migration rate. In general, Wakeley proposes a “scattering phase” (processes of migration and genetic drift within structured populations), which is fast in high – numbered demes (subpopulations). As a consequence the association of variances in the entire population (“collecting phase”) in this case is slow. Thus, in the collecting phase a high number of mutations emerge, while the scattering phase is too short to fix mutations (Wakeley 2001). In many slightly subdivided populations like *D. melanogaster* the populations of the ancestral species range show more variability in genetic and morphology than those in periphery (David & Capi 1988).

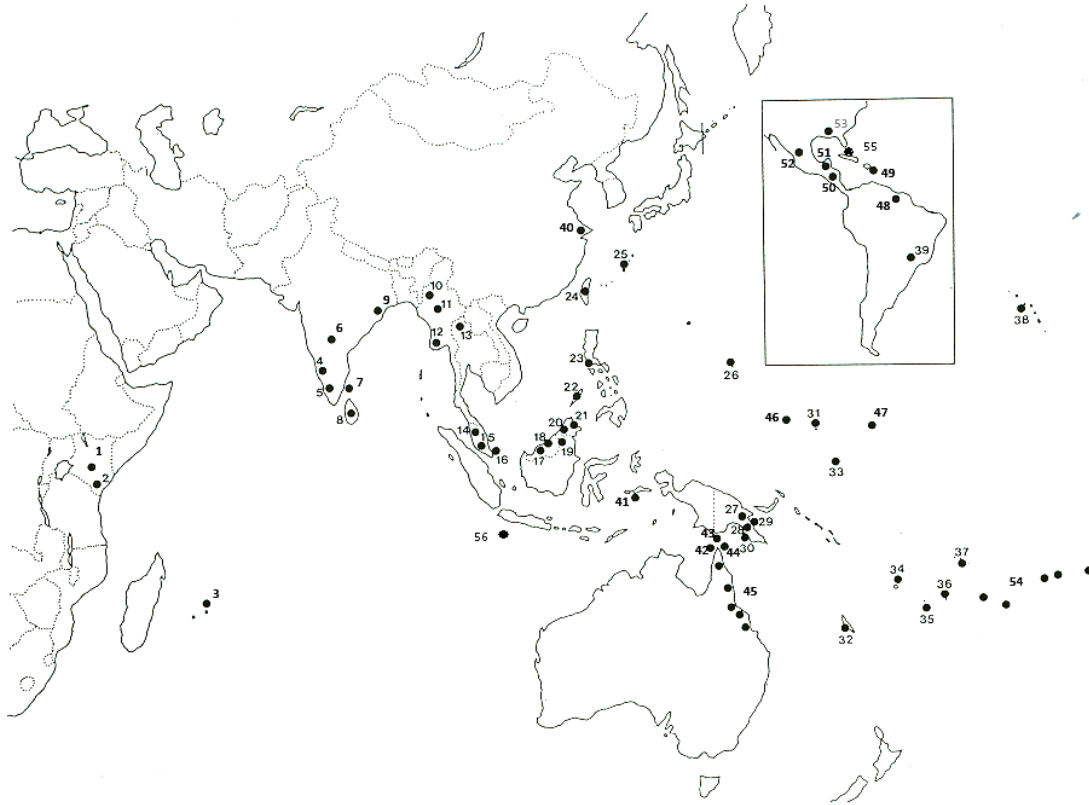


Figure 2 | Distribution of known recorded *D. ananassae* populations:

1 Nairobi, Kenya; 2 Mombasa, Kenya; 3 La Nicoliere, Mauritius; 4 Mysore, India; 5 Coimbatore, India; 6 Hyderabad, India; 7 Madras, India; 8 Kandy, Sri Lanka; 9 Calcutta, India; 10 Shwebo, Myanmar; 11 Mandalay, Myanmar; 12 Yangon, Myanmar; 13 Chiang Mai, Thailand; 14 Penang, Malaysia; 15 Kuala Lumpur, Malaysia; 16 Bukit Timah, Singapore; 17 Kuching, Sarawak, Malaysia; 18 Miri, Sarawak, Malaysia; 19 Brunei Town, Brunei, Borneo; 20 Kota Kinabalu, Sabah, Malaysia; 21 Sandakan, Sabah, Malaysia; 22 Palawan, Philippines; 23 Los Banos, Luzon, Philippines; 24 Taiwan; 25 Nago, Okinawa, Japan; 26 Guam, Marianas Islands; 27 Bulolo, Papua New Guinea; 28 Wau, Papua New Guinea; 29 Lae, Papua New Guinea; 30 Port Moresby, Papua New Guinea; 31 Ponape, Caroline Islands; 36 Noumea, New Caledonia; 33 Nauru; 34 Lautoka, Fiji; 35 Tongatapu, Tonga; 36 Tava'u Islands, Tonga; 37 Pago Pago, Samoa; 38 Honolulu, Hawaii; 39 Ubatuba, Brazil; 40 Shanghai, China; 41 Amahai, Maluku; 42 Badu Island, Queensland, Australia. 43 Mabuig Island, Queensland, Australia; 44 Moa Island, Queensland, Australia; 45 different populations, Queensland, Australia; 46 Colonia, Yap, Micronesia; 47 Marshall Islands; 48 Georgetown, Guyana; 49 Puerto Rico; 50 Nicaragua / Honduras; 51 Campeche, Mexico; 52 Nayarit, Mexico; 53 Louisiana, U.S.A ; 54 several samples from different south pacific islands; 55 Marathon, Florida; 56 Christmas Islands

Adapted from DROSOPHILA ANANASSAE Genetical and Biological Aspects (Tobari 1993). This list is not intended to be exhaustive.

Genetic polymorphism can be detected most frequently in populations that have been stable during a long evolution time within their population size. Frequently, genetic subdivision benefits

from isolation by distance and hence of a limited gene flow in local and long established populations (David & Capy 1988). *D. ananassae* shows only little isolation caused by distance within some Indian populations. Despite that, populations from the original species range are as twice as variable as some Indian populations. So, migration (gene flow) causes a reduction in genetic divergence in due to a shorter “collecting phase” in small populations. Indeed, subpopulations get stabilized by limited genetic variance (Hedrick 2005; Vogl *et al.* 2003; Wakeley 2001).

1.2 Insect’s sense of smell

The sense of smell allows all animals to perceive their surrounding chemical environment. Chemosensory systems, like olfaction are a crucial feature to detect different sources of danger (i.e., toxic substances, contest or scramble competition and contact with natural enemies) and beneficial resources (e.g., food sources, mate selection or proper places for oviposition etc.) via small volatile molecules (Hansson *et al.* 2010; Vosshall & Stocker 2007). Detecting only a waft of a smell is essential for survival respectively finding proper food sources. Like the vinegar fly *D. melanogaster*, *D. ananassae* is found in connection with fruits and vegetables, which constitute the majority part of the food consumed. Fruits produce a huge amount of odorous volatile compounds many of them are detected by drosophilid flies. The aroma volatiles of fresh pineapple is a mixture of at least 280 different chemical compounds whereas only a few of them characterize the typical pineapple odor (Tokitomo 2007; Tokitomo *et al.* 2005). The most abundant volatiles in pineapple are esters such as ethyl hexanoate and methyl hexanoate, which are present in a considerable proportion (Pickenhagen *et al.* 1981).

Flies are able to detect a large amount of odors despite having a “nose” considerably simpler with respect to the vertebrate system. The principal morphological design of the vertebrates and the insects olfactory sensory neurons (OSNs) is however surprisingly similar. Recent comparative studies investigating the morphology and working mode of odorant receptors as well as signal transduction within the chemosensory pathway of olfaction confirmed fundamental differences in vertebrates and insects (Benton *et al.* 2009; Sato *et al.* 2008; Wicher *et al.* 2008). The insect’s sense of smell and its anatomical and neuroanatomical basics are described in the next chapter in more detail.

Anatomy of peripheral organs of the olfactory system in *D. melanogaster*

Olfactory organs of flies are located on the head solely. The olfactory sensory neurons (OSNs) of the fly are housed on the third segment of the antenna as well as on the maxillary palps (see **Figure 3 A and 3 B**). The antenna consists of three segments of which exclusively the third segment (called the funiculus) is covered with multiporous sensilla. These are different in type, size and, morphology. Sensilla types are distributed in a stereotyped way and with a bilateral symmetry on the antenna and can be distinguished in three main types, namely trichoid, coeloconic and basiconic (divided in 3 subtypes: small, thin and large). The different sensilla types detect different types of chemicals in the insect's environment. Basiconic sensilla are known to perceive food odors and CO₂ whereas the trichoids primarily recognize pheromones (Shanbhag *et al.* 1999; Vosshall & Stocker 2007).

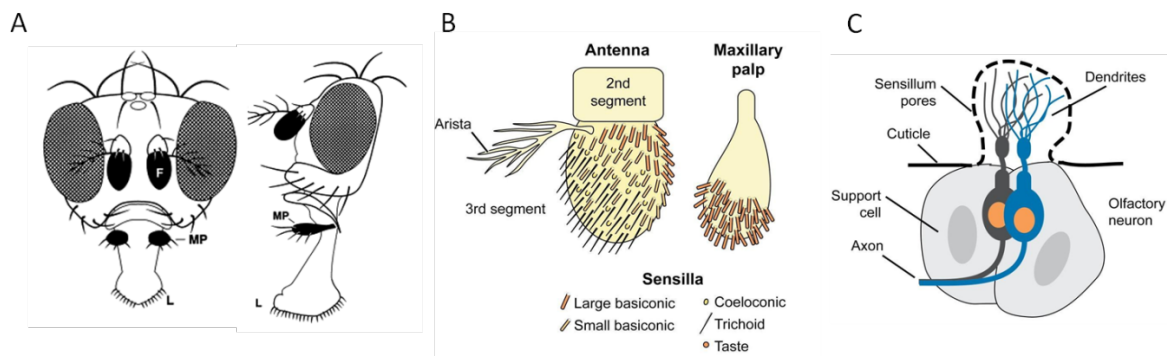


Figure 3 | Anatomy and Neuroanatomy of the peripheral olfaction system of *D. melanogaster*. **A** Schematic drawing of the head of *D. melanogaster*. All olfactory organs like the antenna and maxillary palp depicted in black. Left: frontal view with the antenna in resting position and a half-way extended proboscis. arista (A), funiculus (F), labellum (L), maxillary palp (MP), Figure adapted from (Shanbhag *et al.* 1999); **B** Schematic view of the olfactory organs together with the distribution of sensilla types on the 3rd segment of the antenna and the maxillary palp; **C** Schematic structure of an olfactory sensillum, hosting two ORNs (grey and blue). Adapted from (Vosshall & Stocker 2007)

More recent studies have revealed that coeloconic OSNs express members of a newly discovered gene family of so called ionotropic receptors (IRs) (Benton *et al.* 2009). Chemical stimuli detected by IR expressing OSNs include water vapor, ammonia and acids (Yao *et al.* 2005). The pattern of distribution differs between varying sensillum types. *S. trichodea* are clustered at the lateral – distal edge, large basiconic sensilla at the medial – proximal side, and small basiconic and

coeloconic sensilla in the median region of antenna (see **Figure 3 B**). Another option to classify sensilla types is based on odor response spectra of their corresponding OSNs. Thus, large basiconic sensilla are separated in three different types called ab1, ab2 and ab3 (**b**asiconic sensilla on the antenna, see also (de Bruyne *et al.* 1999; de Bruyne *et al.* 2001).

Each antenna houses between 1100 and 1250 OSNs whereas each single sensillum carries one to four OSNs (see **Figure 3 C**). OSNs within its sensilla type are called e.g., ab3A and ab3B in the case of sensillum ab3 housing two different OSNs A and B (de Bruyne *et al.* 2001). To prevent electrical interference with neighboring cells, each OSN is surrounded by support cells, which isolate the sensillum by producing sensillum lymph. To reach their cognate transmembrane receptor, most of the hydrophobic odorants have to bind to specific odorant binding proteins (OBPs) to transport them in the aqueous environment of the sensillum (Carey& Carlson 2011; Pelosi *et al.* 2006; Vogt& Riddiford 1981). The bipolar structure of OSNs enables a connection between the sensory dendrite ending in the sensillum and, the extended single axon that terminates in the olfactory glomeruli in the antennal lobe (AL) (Vosshall& Stocker 2007).

The AL is the first olfactory neuropil in the fly's brain. The so-called antennal nerve is a bundle of the axons of all OSNs and it extends from the second antennal segment and the arista into the AL attached to the auditory fibers. Thus the AL represents the target for receptor neurons spreading out from the sensilla (Naresh Singh& Nayak 1985). The primary olfactory association center (AL) is structured in 43 morphologically defined glomeruli (Laissue *et al.* 1999). Glomeruli themselves are discrete, spherical structures comprised of densely packed neuropile. Early as well as recent studies have confirmed that OSNs expressing a given OR only target one or two glomeruli – the so-called “one OR to one glomerulus rule” (Couto *et al.* 2005; Gao *et al.* 2000; Vosshall 2000). In the AL projections from all three sensilla types tend to form a clustered pattern (Couto *et al.* 2005). AL with its glomeruli forms the primary centers of olfaction. Cholinergic projection neurons process the information into higher brain centers, the lateral horn and the mushroom bodies. The last mentioned are considered to own a key function in learning and memory (Heisenberg 1998).

1.3 Odorant receptors

ORs in the vinegar fly were first identified in 1999. Although the underlying gene family for expressing an OR was known 8 years before from mammalian studies (Buck& Axel 1991), insect ORs could not be found by using putative sequence homology with mammalian sequence queries

only. Finally, biostatistical approaches and laboratory work exposed a family of seven transmembrane domain proteins, encoded by a large gene family (Clyne *et al.* 1999; Gao & Chess 1999; Vosshall *et al.* 1999). In *Drosophila* 60 OR genes are known so far but unlike ORs of vertebrates, these proteins show no homology to G protein – coupled receptors (GPCRs) (Vosshall *et al.* 1999). On the contrary, biostatistical and also laboratory studies showed that ORs might present a new family of membrane spanning proteins. Regarding its topology the members of this protein family are inverted in comparison with that of GPCRs (Benton *et al.* 2006; Wistrand *et al.* 2006). Via alternative splicing a total number of 62 ORs is received (Robertson *et al.* 2003). Vinegar flies have four different chromosomes (one sex – chromosome and three autosomes) appearing pairwise in each single fly cell, whereas the fourth chromosome is small and does not carry much information. Genes coding for ORs are uniformly distributed all over the three major chromosomes 1, 2 and 3 (1 = sex chromosome, 2 and 3 = the first two autosomes, the third autosome is not named). A few OR genes form small clusters mostly comprise two or three genes that are only distantly related. Only some of them are caused by recent gene duplication. Within the OR gene family the amino acid homology is very low, only about 20%. According to that the absence of large clusters assumes the family of OR genes to be very ancestral (Robertson *et al.* 2003; Vosshall & Stocker 2007).

1.3.1 Expression pattern of OR genes

To detect expression patterns of every single OR gene, various types of methods were used like e.g. RNA in situ hybridization or reporter gene systems like the GAL4:UAS system. The last mentioned tool is used in genetics to express various proteins with a specific driver in a spatial region of tissue only. In the yeast *Saccharomyces cerevisiae* the expression of the GAL4 protein is natively triggered by the presence of galactose. If the GAL4 protein is expressed it binds to an upstream activated sequence (UAS), which induces the transcription of a gene further downstream. The use of the GAL4 system in *D. melanogaster* enables the activation of almost any target gene sequence (e.g., OR gene, green fluorescence protein (GFP) or any other random gene) driven by the UAS. Thereby, the actual target gene and its transcriptional activator are separated from each other in two different transgenic fly lines to guarantee a normal parental development. In the fly line carrying the UAS, the target gene is silent because of the absence of its activator. Activator protein in the GAL4 lines is existent but there is no target gene, which it may activate. If

both lines are crossed to each other, the progeny carry the GAL4 gene and an UAS which triggers the activation of a target gene (Brand & Perrimon 1993) (see **Figure 4**).

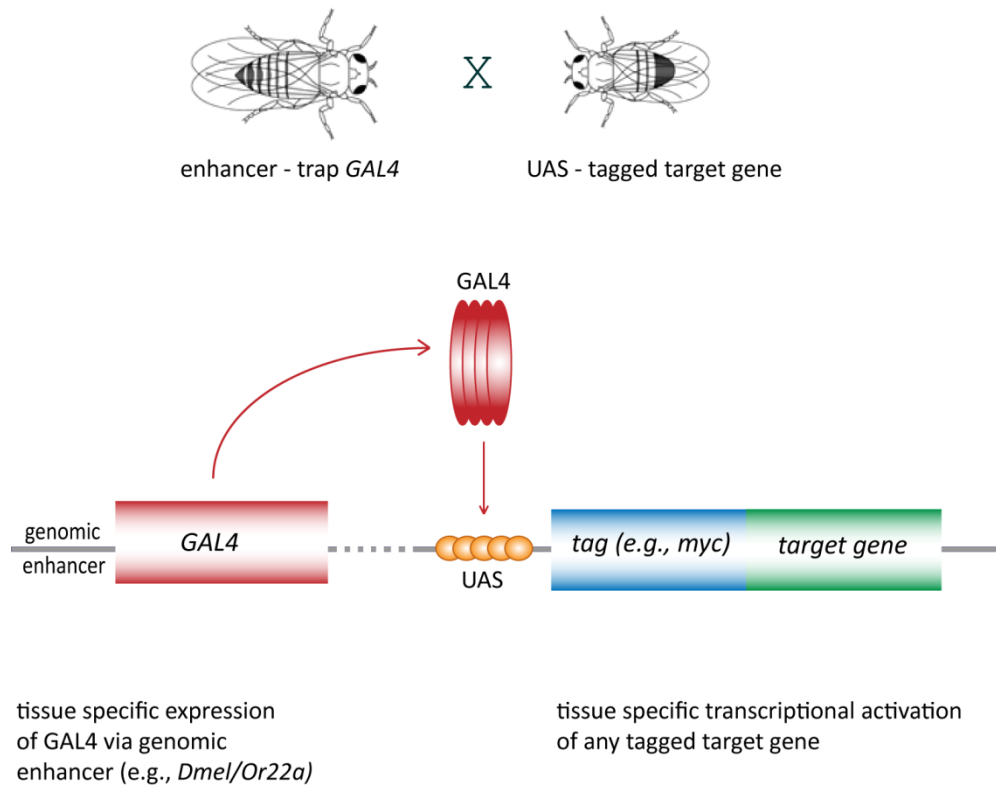


Figure 4 | Schematic drawing about the GAL4:UAS system in transgenic *D. melanogaster*. Two separated transgenic flylines were crossed. One carries a GAL4 sequence which is driven by a genomic enhancer (e.g. *DmelOr22a*). The other flyline hosts a target gene (maybe tagged with myc or GFP) which is driven by an upstream activated sequence (UAS). UAS activation domain consists of 4 GAL4 – binding sites and is only active in progeny generation because of presence of GAL4 protein. System enables expressing of various proteins with a specific driver in a spatial region of tissue. Figure adapted from (Muqit & Feany 2002).

It could be confirmed via RNA in hybridization that 32 OR genes are expressed in the antenna (Vosshall *et al.* 1999). Studies using the GAL4:UAS system indicate that each OR gene is expressed in a spatial conserved subpopulation of OSNs and that only one or a few OR genes are expressed by a single OSN. Solely, Orco (formerly termed *DmelOrco* or *DmelOr83b*) is present in almost all OSNs (Clyne *et al.* 1999; Dobritsa *et al.* 2003; Vosshall 2000; Vosshall *et al.* 1999). Orco associates with the ligand – binding OR in the endomembrane and thus represents a necessary coreceptor in about two third of all OSNs (Benton *et al.* 2006; Larsson *et al.* 2004). By using single sensillum recordings every OR can be assigned to its specific sensillum type. Simultaneously, an

identification of most odorous ligands was intended (Goldman *et al.* 2005; Hallem *et al.* 2006; Kreher *et al.* 2005).

1.3.2 The odorant receptor 22a

The OR 22a detects various odor ligands, which are found in diverse fruit aroma mixtures like these of apricot, cherry, guava, mango, pineapple and plum. The basis for its detection of fruity odors is the underlying ab3A neuron which is excitable to ester ethyl butyrate, ethyl hexanoate and pentyl acetate (found in e.g. banana and apple). The ab3B neuron is sensitive to heptanone and hexanol (de Bruyne *et al.* 2001; Dobritsa *et al.* 2003). Heptanone is a common food additive, which is found in certain foods like beer, bread and potato chips and hexanol is an alcohol produced natively in nature during alcoholic fermentation of rotting fruits. Thus, the OR 22a mediates chemical information to find highly beneficial food sources.

Genetically, *DmelOr22a* forms a cluster with its gene duplicate *DmelOr22b*. They are tandemly arranged and closely related, regarding the nucleotide variation by recent gene duplication (Clark *et al.* 2007; Guo & Kim 2007; McBride & Arguello 2007; Robertson *et al.* 2003). The shared amino acid identity of both genes is comparatively high (78%). It is thought that this gene cluster arose by a recent gene duplication event, whereas *DmelOr22a* itself seems to be responsible for correct odor response in its underlying neuron (Dobritsa *et al.* 2003). Gene duplication events in species are known to be crucial for gaining new functions (Ohno 1970). In 2003 researchers assigned the expression pattern of *DmelOR22a* in the dorso – medial region of the fly antenna. By using the GAL4 system it was revealed that only a few large, basiconic sensilla expresses OR 22a (Dobritsa *et al.* 2003).

The identification of orthologous *Or22a* genes was accomplished using computational approaches based on the *D. melanogaster* genome sequencing project (Adams *et al.* 2000; Kim & Carlson 2002). Finally, two transcripts of *Or22a* were detected in *D. melanogaster*. In addition, 11 other *Drosophila* species were studied to find the origin and diversity of different OR genes as mentioned here for *Or22a*. Especially, the *Drosophila* species *D. ananassae* shows obvious varieties in number of genes, pseudogenes and gene fragments. Unlike *D. melanogaster* *D. ananassae* preserves a total number of six *Or22a* genes within two presumed pseudogenes. *DanaOr22a* genes are tandemly arranged whereas the first and fifth gene seems to be pseudogenes, the second gene differs in length of the third intron. Additionally, the fourth gene

obtained distinct features in sequence in due to alternative splicing (Guo& Kim 2007) (see **Figure 5**).

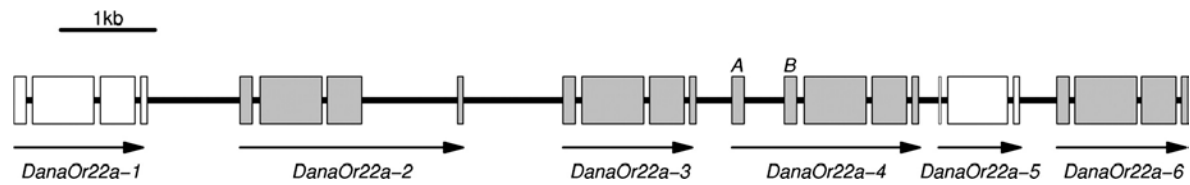


Figure 5 | Tandem array of the 6 *Or22a* genes in *Drosophila ananassae*

White rectangles represent exons in pseudogenes. Grey rectangles represent exons. A functional *DanaOr22a* gene has 4 exons. Arrows indicate gene orientation. In *DanaOr22a-4* alternative splicing was found in due to 2 different forms of the 1st exon, labeled as “A” and “B”. Two first exon isoforms both encode 42 – amino acid segments that differ in 16 positions. Figure adapted from (Guo& Kim 2007)

1.4 Purpose of the Study

The aim of this study was to investigate if all six proposed *DanaOr22a* genes (predicted by Guo & Kim and McBride & Arguello, 2007) are *de facto* existent in the *D. ananassae* genome. Additionally, potential protein properties according to the detected sequences should be analyzed. Differences in *Or22a* genes of all included populations can give rise about the function of differentially expressed ORs. The construction of detailed phylogenetic trees did help to decipher genetic relationships between the different *Or22a* genes of *D. ananassae* as well as the *D. melanogaster Or22a-* and *Or22b* genes. Transmembrane predictions can provide insight into OR protein folding and integration into the OSN membrane. To confirm the raised predictions, sequences were tested *in vivo* via ectopic gene expression using the GAL4:UAS system and subsequent immunostainings.

In theory genes that occurred by recent gene duplication often gain function due to the use of novel ecological niches and thus may provide selective advantages for the animal. This thesis is supposed to yield a comprehensive vantage point for further studies regarding alternative functioning of different *DanaOr22a* genes.

2. Material

2.1 Fly stocks

Table 1 | shows all used flystocks and references about used *D. ananassae* sequence information.

abberation	stock number	locality	country	date collected	donor	company
<i>Dana</i> (QL)	14024-0371.11	Townsville, Queensland (QL)	Australia	–	–	Drosophila Species Stock Center, University of California, San Diego
<i>Dana</i> (FL)	14024-0371.12	Marathon, Monroe County, Florida (FL)	U.S.A.	2002	Birdsley, J.	Drosophila Species Stock Center, University of California, San Diego
<i>Dana</i> (HI)	14024-0371.13	Hawaii (HI)	U.S.A.	1945	Matsuda, M.	Only sequence query used (Clark et al. 2007; Gou& Kim 2007; McBride& Arguello 2007)
<i>Dana</i> (CI)	–	Christmas Island (CI)	Australia	2007	Stensmyr, M.C.	collected during field experiments on Christmas Island

2.2 Bacterial strain

One Shot® TOP10 chemically competent *E. coli* cells

Invitrogen, Karlsruhe/ G

2.3 Cloning vectors

pCR®2.1-TOPO® vector

pUAST

Invitrogen, Karlsruhe/ G
existing in the the molecular laboratories of the Max – Planck – Institute for Chemical Ecology in Jena

2.4 Synthesized Oligonucleotides

All utilized Oligonucleotides were constructed by Eurofins MWG Operon, Ebersberg/ G. **Table 2** shows used primers for the PCR amplification of genomic DNA.

Table 2 | used primers for PCR reaction of genomic DNA templates.

<u>Name</u>	<u>sequence</u>	<u>T_m</u>
DanaOr22aP1.fwd	5'-ATGTTGGGTAGGTTATTTCCCTC	56.5 °C
DanaOr22aP1.rev	5'-TTATTGAAACTTTTCCGCCAG	54.0 °C
DanaOr22aP2.fwd	5'-ATGCTGAGCAAGTTATTTCCGC	58.4 °C
DanaOr22aP3.fwd	5'-ATGTTGAGCAAGTTGTTTCCGAG	58.9 °C
DanaOr22aP3.rev	5'-CTACTGGAAGTTCTCAGCCAG	59.8 °C
DanaOr22aP4.fwd	5'-ATGCTGAGCAAGTTGTTCCCG	59.8 °C
DanaOr22aP5.fwd	5'-ATGTTGAGCAATCTATATCCCCTG	59.3 °C
DanaOr22aP6.fwd	5'-ATGTTGAGTAAGGTATTTCCCTGGATA	60.4 °C
DanaOr22aP6.rev	5'-TCATTGAAATTTTTCGGCCAAGTTCAAT	59.3 °C
MYC_FOR	5'-GAATTCAACATGGAGCAAAAAG	54.0 °C
MYC_REV	5'-AGATCTATTCAGATCCTCCTC	55.9 °C
Or22a2_BglII_fwd	5'-GGAAGATCTATGCTGAGCAAGTTATTTCCGC	66.8 °C
Or22a3_BglII_fwd	5'-GGAAGATCTATGTTGAGCAAGTTGTTTCCGAG	66.9 °C
Or22a6_BamHI_fwd	5'-CGCGGATCCATGTTGAGTAAGGTATTTCCCTGGATA	70.6 °C
Or22a-2rev.McB	5'-CTGAAATTTTTCAGCAATATTGAATTGTCT	58.6 °C

Primers used for colony PCR are listed in **Table 3** stated below.

Table 3 | Used primers for colony PCR reactions.

<u>name</u>	<u>sequence</u>	<u>T_m</u>
M 13 Reverse	5'-CAG GAA ACA GCT ATG ACC	53.7 °C
T 7	5'-TAA TAC GAC TCA CTA TAG GG	53.2 °C

Primers used for sequencing reactions in **Table 4** stated below.

Table 4 | Used primers for all sequencing reactions.

<u>name</u>	<u>sequence</u>	<u>T_m</u>
M 13 reverse (-29)	5'- CAG GAA ACA GCT ATG AC	
T7 Promoter	5'- CCC TAT AGT GAG TCG TAT TA	
myc_for	5'-GAATTCAACATGGAGCAAAAAG	54.0 °C

Or22a_2middle	5'-AACTATTTTCGTGCAATTCCTTCTG	58.1 °C
Or22a_2rev	5'-CCCGATCTAAATAAATAAAAGCG	55.3 °C
Or22a-2rev.McB	5'-CTGAAATTTTTCAGCAATATTGAATTGTCT	58.6 °C
Or22a_3midrev	5'-TTGGCCTCCTGGAACCTCGG	61.4 °C
Or22a_3middle_fw	5'-GAATCGATCCCCGACAGGA	56.0 °C
DanaOr22aP3.rev	5'-CTACTGGAACCTCTCAGCCAG	59.8 °C
Or22a6_middlefw	5'-GACCATCACCTCATCCTGGA	59.4 °C
Or22a6_middlerev	5'-CACCTTTTATTTAGGGGCTCTGTC	61.0 °C
DanaOr22aP6.rev	5'-TCATTGAAATTTTTCGGCCAAGTTCAAT	59.3 °C

2.5 Enzymes

<u>name</u>	<u>distributor</u>
Advantage® 2 DNA Polymerase Mix	Clontech-Takara Bio Europe, Saint-Germain- en-Laye/FR
Alkaline Phosphatase, Calf Intestinal (CIP)	New England BioLabs, Frankfurt a. M./ G
BamHI	New England BioLabs, Frankfurt a. M./ G
BglII	New England BioLabs, Frankfurt a. M./ G
EcoRI	New England BioLabs, Frankfurt a. M./ G
KpnI	New England BioLabs, Frankfurt a. M./ G
Taq DNA Polymerase	QIAGEN, Hilden/ G
T4 DNA Ligase	Invitrogen, Darmstadt/ G
XbaI	New England BioLabs, Frankfurt a. M./ G

2.6 Kits

<u>Name</u>	<u>distributor</u>
DNeasy Blood & Tissue Kit	QIAGEN, Hilden/ G
E.Z.N.A.® Gel Purification Kit	Omega Bio- Tek, VWR International GmbH, Darmstadt/ G
E.Z.N.A.® Plasmid Miniprep Kit II	Omega Bio- Tek, VWR International GmbH, Darmstadt/ G

TOPO TA Cloning Kit
 QIAGEN Plasmid Midi Kit

Invitrogen, Darmstadt/ G
 QIAGEN, Hilden/ G

2.7 DNA ladder/ size marker

2-Log DNA Ladder (0.1–10.0 kb)

New England BioLabs, Frankfurt a. M. / G

2.8 Chemicals and consumables

<u>Name</u>	<u>distributor</u>
Acetic acid	ROTH GmbH & Co, Karlsruhe/ G
Acetone	ROTH GmbH & Co, Karlsruhe/ G
Agar agar	ROTH GmbH & Co, Karlsruhe/ G
Agarose, Seakem LE	Cambrex, Rockland/ ME, USA
Ampicillin	ROTH GmbH & Co, Karlsruhe/ G
Bacto- peptone	ROTH GmbH & Co, Karlsruhe/ G
Bromophenole blue	ROTH GmbH & Co, Karlsruhe/ G
Bacto- tryptone	ROTH GmbH & Co, Karlsruhe/ G
Bacto- yeast extract	ROTH GmbH & Co, Karlsruhe/ G
Diethylpyrocarbonate/DEPC	Sigma – Aldrich, Hamburg/ G
dNTPs, Roti®-Mix PCR 3	ROTH GmbH & Co, Karlsruhe / G
Ethanol	ROTH GmbH & Co, Karlsruhe/ G
Ethidium bromide	ROTH GmbH & Co, Karlsruhe/ G
Ethylenediaminetetraacetic acid (EDTA)	ROTH GmbH & Co, Karlsruhe/ G
Glycerol	ROTH GmbH & Co, Karlsruhe/ G
heat inactivated normal goat serum	Invitrogen, Darmstadt/ G
Isopropyl alcohol	ROTH GmbH & Co, Karlsruhe/ G
MgCl ₂	QIAGEN, Hilden/ G
NaCl	ROTH GmbH & Co, Karlsruhe/ G
Na ₂ HPO ₄ (MW 142) (0.02 M)	ROTH GmbH & Co, Karlsruhe/ G
NaH ₂ PO ₄ (MW 120) (0.03 M)	ROTH GmbH & Co, Karlsruhe/ G
Nuclease free water (DEPC water see below)	

Paraformaldehyde	ROTH GmbH & Co, Karlsruhe/ G
Tissue-Tek® O.C.T™ Compound	SAKURA Finetek Europe B.V., Alphen aan den Rijn, NL
TritonX-100	Sigma – Aldrich, Hamburg/ G
TRIS	ROTH GmbH & Co, Karlsruhe/ G
Vectashield	Vector Laboratories, Inc. Burlingame/ CA/ USA
X-Gal	ROTH GmbH & Co, Karlsruhe/ G

2.9 Antibiotics

Ampicillin (50mg/ml)	ROTH GmbH & Co, Karlsruhe/ G
----------------------	------------------------------

2.10 Buffers and solutions

Gel electrophoresis

6x loading dye (10ml)
 3ml glycerol (100%)
 7ml DI H₂O
 2.5 mg Bromophenole blue

Gel extraction

50x TAE – buffer (1l)
 242g TRIS base
 57,1ml acetic acid
100ml 0.5 M EDTA (pH 8.0)
 Fill up to 1l with DI H₂O

1x TAE – buffer (5l)
 100 ml 50x TAE - buffer
 Fill up to 5l with DI H₂O

Agarose gel (1%)
 100ml 1x TAE-Buffer
 1g agarose
 + 5µl Ethidium bromide [10mg/ml]

6x gel loading dye (10ml)
 2,5g bromophenole blue
 3ml glycerol
 10ml di H₂O

Culture media

All used growth media for *Escherichia coli* were autoclaved for 20 minutes at 121°C and 1.2 bar.

LB - medium (lysogeny broth medium) 1l

10g Bacto-tryptone

10g NaCl

5g Bacto-yeast extract

Fill up to 1l with di H₂O, pH 7.5 (NaOH)

Stored at room temperature

LB_{Amp} – medium (100ml)

LB - medium + 100µl ampicillin

Stored at 4°C

LB_{Amp}-Agar (100ml)

LB - medium + 1.5 g Agar agar

To dissolve the Agar agar the mixture had to be cooked in a microwave.

+ 60µl ampicillin [50mg/ml]

poured in petri plates (10 cm diameter)

stored at 4°C

S.O.C. – Medium for competent *Escherichia coli* cells

The S.O.C. Medium was enclosed to the chemical competent *E. coli* kit (One Shot®TOP10 kit, Invitrogen).

2.11 Immunostaining

Buffers and solutions

10 x PBS stock (1l)(washing solution)

NaCl 76 g (1.3 M)

Na₂HPO₄ (MW 142) 9.94 g (0.02 M)

NaH₂PO₄ (MW 120) 3.6 g (0.03 M)

for 1 x PBS buffer 100 ml 10 x PBS were diluted in 900 ml di H₂O.

paraformaldehyde / PBS buffer (fix solution)

4% paraformaldehyde diluted in 1 l 1 x PBS

should be fresh, less than 7 days old

PT buffer (permeability buffer)

1 l 1x PBS with 0.1 % TritonX-100 (Sigma – Aldrich)

PTS (blocking solution)

PT buffer with 5 % heat inactivated normal goat serum (Invitrogen)

AntibodiesPrimary antibodies

α – OR22a (rabbit)

kindly provided by Leslie B. Vosshall

α – c -Myc antibody (mouse)

abcam, Cambridge/ UK

Secondary antibody

α — mouse Alexa Fluor 488 goat

Invitrogen, Darmstadt/ G

α — rabbit Alexa Fluor 546 goat

Invitrogen, Darmstadt/ G

2.12 Laboratory equipment

Besides the general laboratory equipment, following utensils were used:

Centrifuge Type 5810	Eppendorf, Hamburg/ G
Comfort Thermomixer 1.5 ml and 2 ml	Eppendorf, Hamburg/ G
Cryo-Star HM 560 Cryostat MICROM International	Thermo Fisher Scientific, Walldorf/ G
Electrophoresis system, Mupid-exU	Advance, Potsdam/ G
Gel documentation, Bio - Vision	Peqlab, Erlangen/ G
Incubator, Kendro B12 FunctionLine	Heraeus Instruments, Hanau/ G
Mikroskop LSM 510 META	Carl-Zeiss GmbH, Jena / G
Rotary incubator	HT, Bottmigen/ CH
Spectrometer, BioPhotometer	Eppendorf, Hamburg/ G
Thermal cycler, GeneAmp PCR System 9700	Applied Biosystems, Darmstadt/ G

Software

BOXSHADE 3.21	http://www.ch.embnet.org/ software/BOX_form.html / CH
ClustalX2 (Clustal: Multiple Sequence Alignment)	Cornway Institute UCD Dublin/ IRL
FlyBase BLAST Search	http://flybase.org/blast/
MUSCLE (M ultiple S equence C omparison by L og- E xpectation), (Edgar 2004a, b)	
NCBI BLAST Search	NCBI, Bethesda/ USA
PAUP* 4.0	(Swofford 2002)
Seqbuilder	DNASTAR Inc. , Madison/ WI, USA
SeqMan Pro	DNASTAR Inc. , Madison/ WI, USA
Sequencher 4.7 software	GeneCodes, Ann Arbor/ MI, USA

3. Methods

3.1 DNA extraction

Total genomic DNA extraction was performed by using the kit: Purification of total DNA from insects using the DNeasy® Blood & Tissue Kit (QIAGEN) by following the manufacturer instructions. The extracted DNA was stored at -20°C in the Department of evolutionary neuroethology (molecular laboratories) of the MPI for Chemical Ecology in Jena.

Quantity and quality of DNA

The concentration and the purity of the DNA samples were verified by gel electrophoresis and with an optical density measurement on a photo spectrometer.

To check the concentration of the samples 2µl of DNA were diluted in 78µl distilled water and measured at 260nm wavelength. An OD260nm of 1 corresponds to 50µg/ml double stranded DNA.

$$c = A_{260} \times F \times K$$

- c: molar concentration [µg/ml]
- A260: measured absorption at a value of 260 nm
- F: dilution factor
- K: concentration of dsDNA (50µg/ml)

To evaluate the purity of the samples the values at 230nm and 280nm were additionally measured. The ratio between A260nm and A280nm provides an estimate of contamination with proteins whereas proteins absorb at 280nm. Pure preparations of DNA should have an A260nm/A280nm ratio of 1.8 to 2. The OD260/230 ratio indicates contamination with organic compounds such as polypeptides. A ratio around 2 indicates a pure DNA sample.

DNA samples were also examined via gel electrophoresis (see **3.4.1** for descriptions).

3.2 Primer design

Primers were designed based on *DanaOr22a* genes sequence data of studies from (Guo& Kim 2007; McBride& Arguello 2007). To guarantee adequate specificity, primers contain about 20 base pairs (bp) length with GC content between 40 – 60 %. The melting temperature of the primers was between at least 55 °C till 70 °C to obtain best results. Given sequences were used as a template to design efficient primers. The first 20 bases of the 5' end were used as forward primers. 20

bases of the 3`end were used as reverse primers. The reverse primer sequence had to be reverse and complementary. In addition a restriction site for different enzymes was added to the forward primers, which later enables to cut the PCR product out of the cloning vector. All primers are listed in **Table 2**. Primers were synthesized by MWG-Biotech (Ebersberg, Germany) and diluted to get a concentration of 10pmol/ μ l prior PCR usage.

3.3 PCR procedure

3.3.1 General mode of action

PCR was used for selective DNA amplification of diverse *Or22a* genes of *D. ananassae*. Genomic DNA from different *ananassae* populations as well as plasmids of transformed *E. coli* cells were used as templates. The DNA was heated up to 94°C to denaturate double DNA strand (dsDNA). At temperature lowered to 50 – 65°C the two oligonucleotides bind their complementary sequences of the denatured single strand DNA (ssDNA) (annealing). The oligonucleotides serve as primers for the thermo stable PCR polymerase which extends primers and fills up the ssDNA to dsDNA (elongation). An additional denaturation of the doubled DNA fragment enables again annealing of the primers, so that the process can be repeated cyclically. Finally specific DNA region accumulates exponentially (Mullis *et al.* ; Saiki *et al.* 1988).

3.3.2 Master Mix and Polymerases

Because of the individual primer design, all primers had different annealing temperatures. The optimal annealing temperature of every primer pair (forward and reverse) is 5°C less than its ordinary melting temperature.

In PCR reactions using Taq DNA Polymerase (Qiagen) each 25 μ l Master Mix contains the following components:

- 18.0 μ l PCR grade H₂O
- 2.5 μ l 10x Taq DNA polymerase Buffer
- 1.0 μ l forward primer [10pmol/ μ l]
- 1.0 μ l reverse primer [10pmol/ μ l]
- 1.0 μ l DNA template [~ 100ng]
- 0.5 μ l MgCl₂ [25mM]

0.5µl 50x dNTP Mix (10mM of dATP, dCTP, dGTP, dTTP)

0.5µl 50x Taq DNA polymerase

25.0µl total reaction volume

In PCR reactions using Advantage® 2 DNA Polymerase Mix (Clontech) each 25µl Master Mix contains the following components:

18.5µl PCR grade H₂O

2.5µl 10x Advantage® 2 DNA polymerase Mix Buffer

1.0µl forward primer [10pmol/µl]

1.0µl reverse primer [10pmol/µl]

1.0µl DNA template [~ 100ng]

0.5µl 50x dNTP Mix (10mM of dATP, dCTP, dGTP, dTTP)

0.5µl 50x Advantage® 2 DNA polymerase Mix

25.0µl total reaction volume

3.3.3 PCR conditions

94°C	3 min	} 37 cycles
94°C	30 sec	
55 – 60°C	1 min	
68°C	1 min	
68°C	10 min	
4°C	∞	

3.4 Agarose gel electrophoresis

3.4.1 General mode of action

The results of DNA extraction and PCR products were verified by using agarose gel electrophoresis. Substances were loaded on a gel and separated in bp length within an electrical field. Short molecules pass the pores of the gel more easily than long ones, which need more time

to run through the gel matrix. Due to the phosphate group of the nucleic acid and its negative electrical charge, DNA move through the gel matrix toward the anode.

To prepare a 1.5% gel 1.5g agarose was dissolved in 100ml TAE – buffer while heated in a microwave. After the solution had been cleared 5 μ l Ethidium bromide (10mg/ml) was added. Ethidium bromide (EtBr) intercalates into DNA strands and thus changes the absorption spectra of DNA, which makes it useful as a fluorescent tag. EtBr induces fluorescence of the nucleic acids in the gel when exposing the gel to ultraviolet light. 5 μ l of DNA or PCR products were first mixed with 2 μ l loading dye. This loading dye contains Glycerol to weigh down the samples to the bottom of the slots. Additionally the dye contains a color marker (Bromophenol blue) to control the process of the electrophoresis. Simultaneously 5 μ l of a DNA marker (2 Log DNA ladder) were loaded on the gel to estimate the length of the unknown fragments or products, respectively. The separation was executed at 135V for 25 minutes. The pattern of DNA-bands were estimated and photographed under UV-light.

3.4.2 Separation and purification of PCR products

PCR products were separated using agarose gel electrophoresis. Samples were loaded on a 1 % agarose gel and separated at 135 V for 25 minutes. The band of the expected size was cut off the gel with an ethanol cleaned scalpel and transferred into a clean tube. All PCR products were purified using E.Z.N.A.[®] Gel Extraction Kit (Q-Spin Column) and eluted with 30 μ l water.

3.5 Transformation and selection

To clone the purified PCR products TOPO TA cloning kit was used by following the manufacturer's protocol. This kit includes a pCR2.1 TOPO vector in which PCR products were cloned (Invitrogen, **Figure 6**) by using the enzyme topoisomerase I. This enzyme has two important functions: first it is a restriction enzyme and second it is a ligase. By recognizing a pentameric sequence, topoisomerase I forms a covalent bond at the 3'thymidin attached phosphate group. To unwind the DNA, topoisomerase I cleaves one DNA strand, allows the DNA to unwind and then religates the ends. To enable the immediate ligation of sequences with compatible ends with the vectors, TOPO vectors are provided linearized. Additionally a topoisomerase I is covalently bound on each

3' phosphate of the vector. Due to these features, the ligation can take place at room temperature and is complete after 5 minutes (Stivers *et al.* 1994) (**Figure 6**).

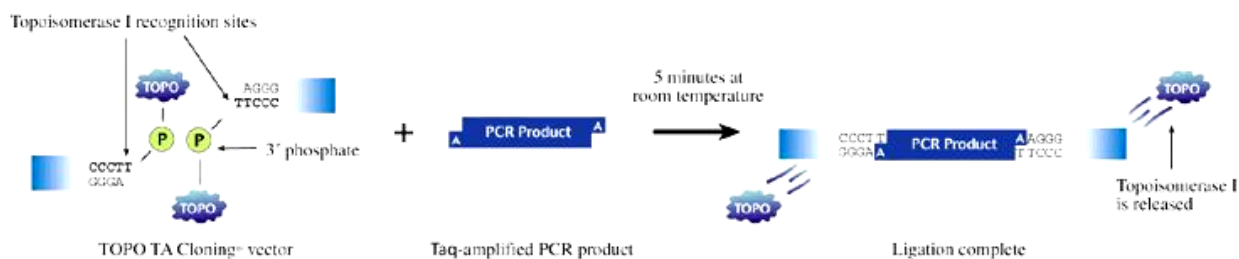


Figure 6 | TOPO® TA Cloning® of Taq-amplified DNA. By the help of the enzyme DNA topoisomerase I, DNA will cleave and rejoin during replication. Topoisomerase I, thereby recognized a pentameric sequence (5'-(C/T)CCTT-3') and enables a promptly ligation of cloning vector and amplified PCR products with compatible ends. In 5 minutes in room temperature the ligation is complete. Adapted from Invitrogen, TOPO TA cloning® kit.

The cloning reaction was performed following the manufacturer's protocol and had a total reaction volume of 3µl and contained the following components:

- 2.0µl purified PCR product
- 0.5µl salt solution
- 0.5µl pCR®2.1-TOPO® vector

For multiplication the 2µl of the ligated plasmids were transformed into 25µl One Shot® TOP10 chemical competent *E.coli* cells (Invitrogen) using heat shock. After 45s at 42°C the transformed cells were mixed with 150µl S.O.C. medium (provided in the TOPO TA cloning kit). After one hour shaking at 37°C, the transformed cells were plated out on LB_{Amp}-Agar plates and grew over night at 37°C.

Selection of transformed *E. coli* strains on antibiotic resistance and blue white screen
 The pCR®2.1-TOPO® vector contains an ampicillin resistance gene. To verify a successful transformation of the plasmid into *E. coli* cells, they were grown in the presence of ampicillin to make sure that just bacteria with an ampicillin resistance survive. Besides the selection about the antibiotic resistance, the pCR®2.1-TOPO® (**Figure 7**) contains a *lacZα* gene with an internal multi

cloning site (MCS). PCR products can be integrated within the MCS of *lacZα* genes by using diverse restriction enzymes. In case of a successful insertion of foreign DNA into the MCS, the production

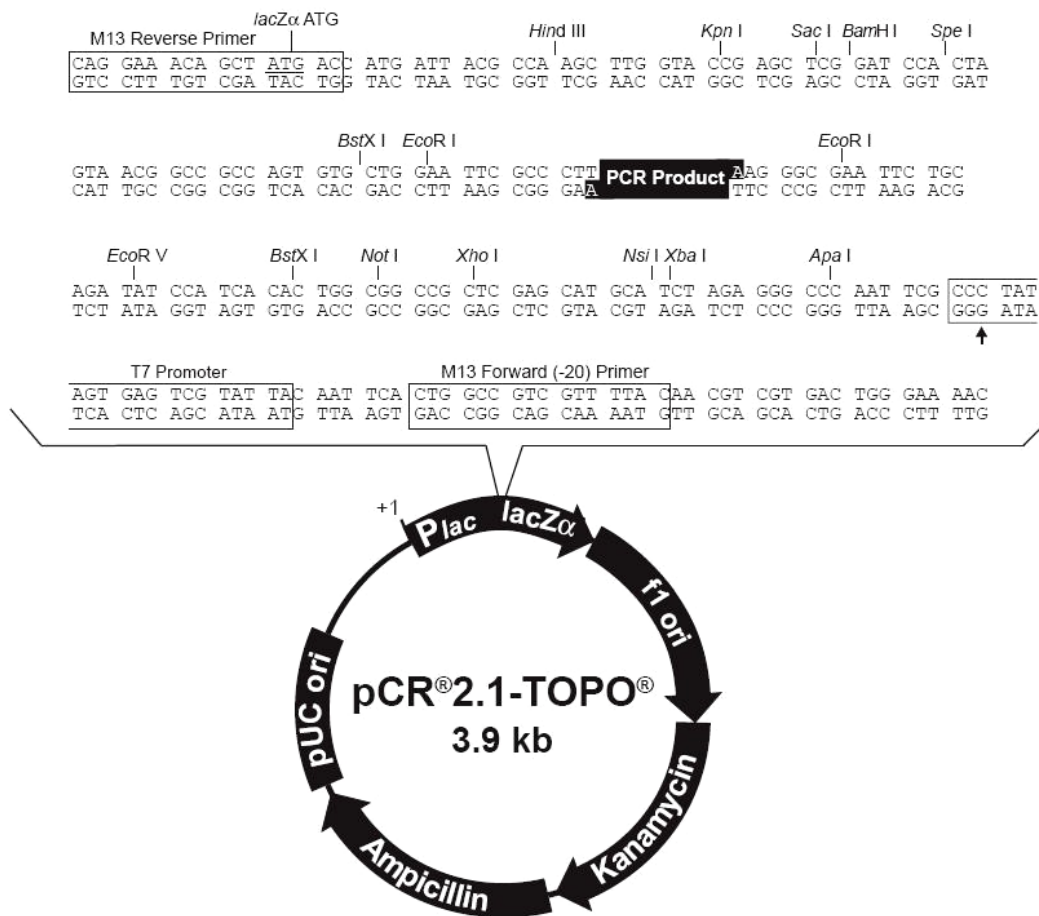


Figure 7 | Map of the pCR[®]2.1-TOPO[®] vector. Vector contains antibiotic resistance for ampicillin and kanamycin and a through Topoisomerase I recognizing pentameric sequence for ligation (e.g., any PCR product). In addition the plasmid encloses different restriction sites for several restriction enzymes. Figure adapted from Invitrogen.

of β – galactosidase is interrupted. Thus, X – Gal was used as an indicator of a successful insertion. The colorless and modified galactose sugar X – Gal is hydrolyzed by β – galactosidase and forms a blue insoluble product in the colonies. Blue colonies produce β -Gal, which indicates they contain vectors without inserts. White colored colonies indicate that PCR products are present in the MCS of *lacZα* gene in pCR[®]2.1-TOPO[®] vector and thus interrupt the ability to hydrolyze the X – Gal marker.

3.6 Colony PCR

To test if picked *E. coli* colonies host the right insert, colony PCR was performed. From every Petri dish several white colonies were picked, precultured in 50µl LB_{Amp}⁻ medium and shaken for one hour at 37°C. A master mix containing the following contents was used:

18.0µl	H ₂ O (PCR grade)
2.5µl	10x PCR buffer
1.0µl	M13 reverse primer [10pmol/µl]
1.0µl	T7 primer [10pmol/µl]
2.0µl	preculture
0.5µl	50x dNTP's (10mM of dATP, dCTP, dGTP, dTTP)
0.5µl	Taq DNA Polymerase
<hr/>	
25.5µl	total reaction volume

The used PCR thermo cycler was programmed with appropriate temperatures and cycles given below:

94°C	2 min	} 35 cycles
94°C	15 sec	
50°C	2 min	
72°C	30 sec	
72°C	5 min	
4°C	∞	

To verify the results, 5µl of the PCR products were separated by size on a 1.5% agarose gel. Precultures of clones which were tested positively were cultured in 5ml (Mini) or 50ml (MIDI) LB_{Amp}⁻ medium over night in a rotary incubator at 37°C.

3.7 Plasmid isolation

To isolate plasmids, the liquid over night culture was centrifuged at 4000rpm for 15 minutes. Plasmid DNA was extracted and purified from the medium by using the E.Z.N.A.[®] Plasmid Miniprep Kit I and eluted once in a volume of 40µl. For MIDI preps QIAGEN Plasmid Midi Kit was used and

eluted in a volume of 100µl TE buffer. For all procedures the manufacturer`s instructions were followed.

The concentration and quality of isolated plasmids was checked via optical density measurement on a photo spectrometer.

3.8 Sequencing and sequence analyses

3.8.1 Sequencing

All plasmid DNA samples were sequenced by MWG-Biotech (Ebersberg, Germany). Therefore, 1µg of plasmid DNA in a total volume of 15µl (fill up with H₂O if necessary) was sent in a 1.5ml tube to the company. M13rev (-29) and T7 promoter primers were used to get both directions of the DNA strand sequenced (for primer sequences see **Table 3**).

For sequencing *Or22a* full length gene additional sequencing primers were used (see **Table 4**).

3.8.2 Sequence analyses

Contig

All sequenced files were edited with Sequencher 4.7 software (GeneCodes). Both clone sequences (M13 and T7) were arranged together to one contig sequence. If clone sequences are ambiguous, Sequencher 4.7 software uses Wobble – bases for the contig sequence as determined in IUB Base Code Guide (Valid International Union of Biochemistry). If bases of M13 and T7 sequences are identically, software uses unambiguous bases (A, C, G and T). All contigs were transformed into the FASTA format for further studies (Pearson& Lipman 1988).

BLAST

A Basic local alignment search tool (BLAST) was used to prove obtained contigs (Altschul et al. 1990; Altschul et al. 1997). Two different BLAST searches were done. The National Center for Biotechnology Information (NCBI, <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and FlyBase.org BLAST search (<http://flybase.org/blast/>) were performed by uploading the sequences to the homepages. BLAST algorithm search enables a comparison between obtained DNA or protein sequences with an existing database. The search is determined by different values such as score or E – value. The score gives a quantitative statement about similarity of the query sequence and the sequence of

interest. The probability that in the database is another alignment with a high similarity occurred by chance and not by descent of the species is given by the E – value (Hall 2011).

Alignment

All sequences with an expected BLAST result were aligned by using the Kim and McBride sequences as template. All alignments were done by using the ClustalX2 program. For “Pretty Printing and Shading of Multiple-Alignment files” ClustalX2 alignment output (ALN file) was pasted into the BOXSHADE3.21 panel (http://www.ch.embnet.org/software/BOX_form.html). As output format on the BOXSHADE3.21 server “RTF new (using shaded background)” was used.

3.8.3 Open reading frame (ORF) and translation into amino acid sequence

Appropriate sequences were checked in Seqbuilder (DNASTAR Inc., Madison/ WI, USA) software to check open reading frames. Therefore, all introns in genomic DNA sequences had to flip out manual. Sequences which were used for further studies should reach a common accord to the template sequence. If that is the case, sequences were translated into amino acid sequence by using the Seqbuilder (DNASTAR Inc., Madison/ WI, USA) software. All amino acid sequences were revised again with its translated template sequences (Guo& Kim 2007; McBride& Arguello 2007) by using an alignment (via ClustalX2 program). Only sequences with a highly identicalness with template sequence, no base deletions and as few as possible changed amino acids were used for making a transgenic fly.

3.9 Phylogenetic analyses

Pylogenetic tree was designed based on given AA sequence data from 4 *D. ananassae* populations. An alignment was made using MUSCLE (**M**Ultiple **S**equence **C**omparison by **L**og-**E**xpectation) (for detailed information see (Edgar 2004a, b)). The alignment includes 15 different sequences. Nine different sequences of *D. ananassae* (four populations each including three *Or22a* genes) and two sequences of ORs of *D. melanogaster* (*DmelOr22a* and *DmelOr22b*, sequence data given by flybase.org). Additionally, the AA sequence of Orco (data from flybase.org) was added into the alignment. The resulting alignment file was used to generate a tree based on maximum parsimony analysis. Parsimony based analyses creates trees, which arrange the dataset by the minimum number of AA changes within the alignment. Trees with a minimum number of

parsimony – informative site steps are selected by the program. Orco was defined to be outgroup to create a rooted tree. Evolutionary tree was designed based on alignment from MUSCLE using PAUP* 4.0 (Swofford 2002).

3.10 Transmembran domain prediction

AA sequences of *DanaOr22a* genes of all 4 populations were used for transmembrane domain prediction analyses. Following tools were used: SPLIT (Juretic *et al.* 1993) and TMpred (Hofmann & Stoffel 1993). These bioinformatical tools are based on mathematical probability calculations. Known positions of potential transmembrane domains of AA sequences were used to create a transmembrane protein topology via TOPO2 (Johns 2005).

3.11 Construct design for making a transgenic fly

A vector named pUAST was used as a base for creating a construct. pUAST vector is about 9050 bp long and it has 5 GAL4 binding sites. Additionally it hosts a polylinker which is situated between the downstream GAL4 binding sites and the SV40 small t intron and polyadenylation site upstream (Brand & Perrimon 1993). In addition an ampicillin resistance is present. MCS or also called the polylinker of the vector contains restriction sites for several enzymes, for example: EcoRI, BglII, KpnI and XbaI (see **Figure 8**).

Additionally, a Myc tag was ordered based on the c-Myc sequence from Sigma – Aldrich, Hamburg/ G. The myc sequence was flanked with an EcoRI restriction site on the 5' end and a BglII restriction site on the 3' end. Synthesized and cloned into a pCR[®]2.1-TOPO[®] vector by MWG-Biotech (Ebersberg, Germany) 10 µg of the c-Myc – tag was sent to MPI for Chemical Ecology (Jena, G). Simultaneously, a pUAST and pCR[®]2.1-TOPO[®] - myc vector were double digested to ligate the myc – tag into the pUAST vector.

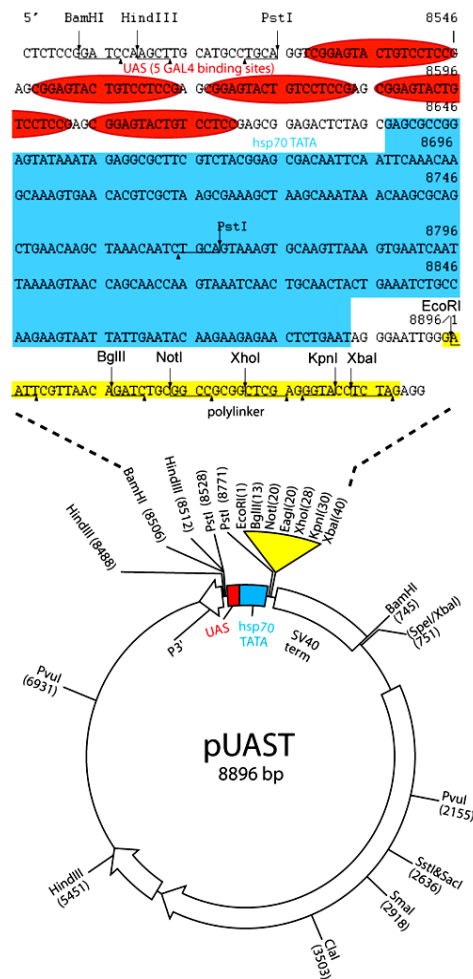


Figure 8 | Map of the pUAST vector to construct a UAS reporter. Target genes were subcloned into the polylinker (GAL4 UAS) of the pUAST Vector. Furthermore, the vector carried an antibiotic resistance (ampicillin) and several restriction sites for enzymes. Figure adapted from (Phelps & Brand 1998).

2µg pUAST	10µg pCR® 2.1-TOPO® - myc vector
10µl NEBuffer 2	3µl NEBuffer 2
2µl BglII	1µl BglII
<u>2µl EcoRI</u>	<u>1µl EcoRI</u>
fill up to 100µl with DI H ₂ O	fill up to 30µl with DI H ₂ O

All digestions were incubated over night at 37°C. After incubation a CIP (Calf Intestine Alkaline Phosphates, NEB) treatment was done. Alkaline phosphatase removes phosphate groups from

proteins in a process called hydrolysis, so it is a hydrolase enzyme. In alkaline environment alkaline phosphatases are most effective. Therefore 1 μ l CIP was added to the reaction and incubated for half an hour at 37°C. The whole volume of the reaction was loaded on a 1% agarose gel to separate the cut vectors. Subsequently, fractionated bands of the expected size were cut out of the gel. Gel slices were extracted and cleaned as already described (see 3.4.2 for descriptions). 2 μ l of cleaned up vectors were loaded on a gel to estimate the quantity. For the ligation charges the following components were used:

<u>Ligation</u>	<u>Autoligation</u>
10 μ l cleaned myc (cut)	
5 μ l aqua dest	15 μ l aqua dest
2 μ l cleaned pUAST (cut)	2 μ l cleaned pUAST (cut)
2 μ l 5X DNA Ligase Reaction Buffer	5X DNA Ligase Reaction Buffer
<u>1 μl T4 DNA Ligase</u>	<u>1 μl T4 DNA Ligase</u>
Σ 20 μ l	Σ 20 μ l

Ligation and autoligation were incubated over night at 16°C. 10 μ l of the ligation were transformed into One Shot® TOP10 chemically competent *E. coli* cells (Invitrogen) and incubated at LB_{Amp}-Agar plates over night at 37°C. All grew colonies were checked for the right insert by colony PCR and gene specific primers.

Colonies which were tested positive were sent out to sequencing with gene and vector specific primers (see **Table 4**).

Into the polylinker of the myc tag housing pUAST vector, different *Or22a* genes were cloned. Therefore all genes as well as the pUASTmyc vector had to be digested with diverse restriction enzymes. All genes were first cut out of the pCR®2.1-TOPO® vector. All contain a BglIII restriction site on their 5' end. *DanaOr22a-2* and *DanaOr22a-6* were double digested with BglIII and XbaI, *Or22a-3* with BglIII and KpnI. Digestion, ligation, transformation, colony PCR and sequencing were done as mentioned above. All complete constructs comprise a pUAST vector containing an N-terminal c – Myc tag and an adjacent full length gene of *DanaOr22a* within the MCS. All constructs were extracted from One Shot® TOP10 chemically competent *E. coli* cells (Invitrogen) by using a E.Z.N.A.® Plasmid Miniprep Kit II. Columns and were eluted with 100 μ l ddH₂O.

3.12 Transgenic fly lines

By the help of the UAS:GAL4 system (see **1.3.1**) it is possible to activate gene expression of *DanaOr22a* gene transcripts in a transgenic *D. melanogaster*. To generate the transgenic *D. melanogaster*, a total of 10µg plasmid DNA per construct has been sent to Sang Chan (Dept. Genetics Cambridge University Downing Street Cambridge CB2 3EH). Subsequent embryo injection, transformant selection as well as balancing of the now transgenic flies were performed in the Sang Chan laboratories.

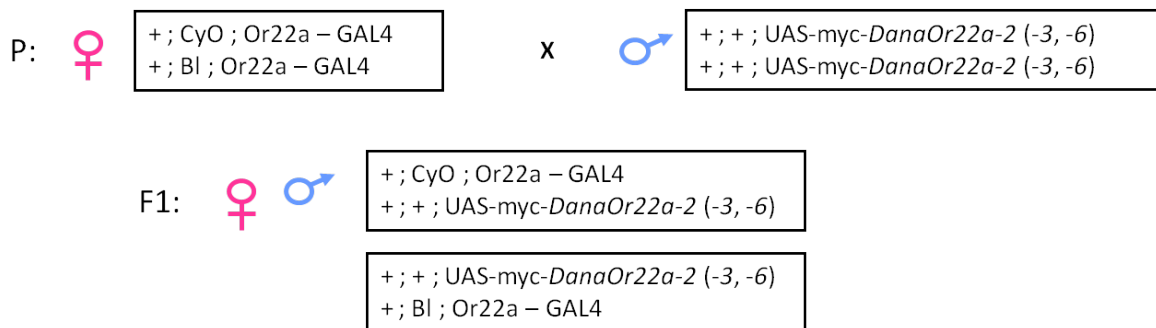


Figure 9 | Schematic picture of the crossing scheme for tissue specific expression of *DanaOr22a* genes in transgenic *D. melanogaster* using UAS:GAL4 system. P parental generation: female carries the tissue specific driver *Or22a* homozygous on the third chromosome. Male carries UAS-myc-*DanaOr22a* of the 3 different odorant receptor genes of *D. ananassae* (-2, -3 and 6). F1 progeny generation: Every fly carries the UAS:GAL4. Different *DanaOr22a* genes emerge additional to the proper *DmelOr22a* in the final transgenic *D. melanogaster* (F1).

Flies received were kept in vials with food in an incubator at 25°C and 70% humidity. Additionally incubator featured 12 hours day and night cycle. Flies were flipped every second week into a new vial with new food. Fly food (1 liter) consisted of 918ml water, 118g sugar beet molasses, 95g polenta, 11g brewer's yeast, 4.1g agarose, 2.4ml propionic acid and 3.3ml nipagine (16%). On their third chromosome transgenic UAS fly lines host a transcript of *DanaOr22a* gene flanked by a myc – tag and additionally an UAS (see **Figure 4**). In this case transgenic GAL4 OR22a lines were crossed with transgenic UAS-myc-*DanaOr22a* lines. Progeny generation carries supposed *DanaOr22a* genes in addition to its proper *DmelOr22a* (see **Figure 9**). To examine right cloning, successful assembly of the foreign genes and expression of OR proteins in the transgenic *D. melanogaster* immunofluorescence was done.

3.13 Immunostaining

Collected progeny flies were embedded in Tissue-Tek® O.C.T™ Compound and sliced in cryostat (MICROM HM 560) afterwards. Longitudinal sections were collected on Fisher Superfrost Plus slides and fixed immediately after sectioning for 7 minutes in paraformaldehyde / PBS buffer. Then sections were washed two times for 10 minutes in 1 x PBS. Permeabilize sections 30 minutes in PT buffer. For 30 minutes each slide was blocked with PTS buffer while placing them into with PT buffer humidified chamber. 200 µl PTS per slide were used. An appropriate dilution of used primary antibody in PTS was prepared (α - OR22a 1:1000, α -c – Myc 1:1000). After removing blocking solution from every slide, 100 µl of each diluted antibody were applied. To prevent evaporation cover slips were added and slides were incubated overnight at 4°C.

After incubation sections were washed three times 10 minutes in PT buffer. Slides were blocked again for 30 minutes in PTS buffer as described above. Secondary antibodies were diluted 1:250 in PTS. 100 µl of dilution were added on each slide incubated at 25°C and shielded from light. After an incubation time of two hours sections were washed three times 10 minutes in PT buffer. Washed slides were mounted in 60 µl Vectoshield and sealed with cover slips. Stored in the dark at 4°C fluorescent signal usually fades in one to four weeks.

Slides were measured with a confocal scanning laser microscope (Mikroskop LSM 510 META, Carl-Zeiss GmbH, Jena / G) and with a wavelength of 488 nm and 543 nm. Different filters were used (Ch2: BP 505 – 550 and Ch3: BP 560 – 615).

4. Results & Discussion

4.1 DNA Extraction

Total genomic DNA was extracted from two taxa *Dana*(QL) and *Dana*(FL). Genomic DNA from *Dana*(CI) did already exist in the molecular laboratories of the Max – Planck – Institute for Chemical Ecology in Jena. The concentration of all DNA samples was checked via optical density measurement (**Table 5**). Additionally, total DNA of all populations was verified by gel electrophoresis. If DNA is of high quality, bands were clear and hardly smeared. **Figure 10** shows successful isolated genomic DNA on a 1.5% agarose gel.

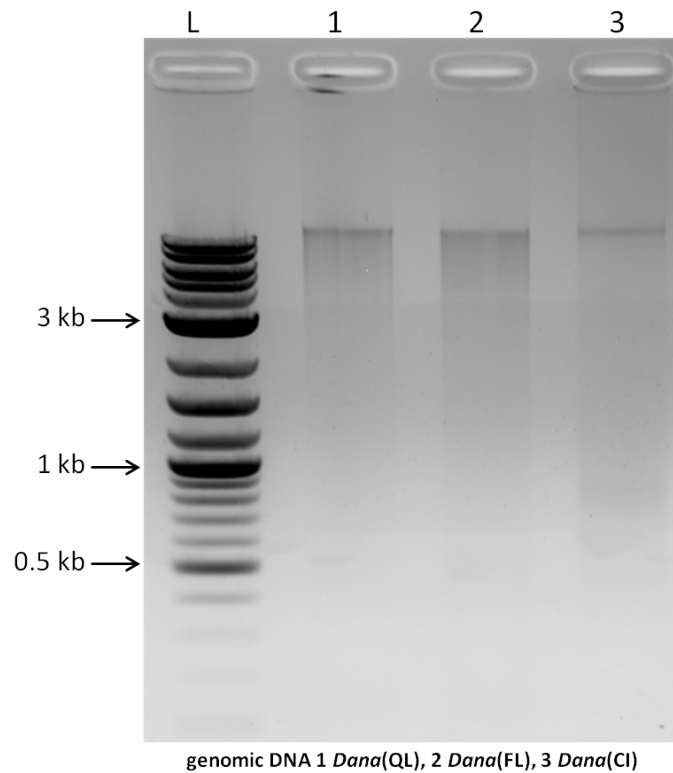


Figure 10 | Pictured extracted DNA of different *D. ananassae* populations. Resulting genomic DNA verified by gelectrophoresis from 3 different *D. ananassae* populations. **1** *Dana*(QL), **2** *Dana*(FL) and **3** *Dana*(CI). Hardly smeared and clear bands indicate good quality of extracted DNA. L 2-Log DNA Ladder.

Table 5| Concentration of genomic DNA of 3 *D. ananassae* populations. Genomic DNA of populations of *D. ananassae* from Queensland, Florida and Christmas Island was extracted and measured via optical density measurement. Concentration of extracted DNA is given in $\mu\text{g/ml}$.

	<i>Dana</i> (QL)	<i>Dana</i> (FL)	<i>Dana</i> (CI)
genomic DNA in $\mu\text{g/ml}$	68,3	25,8	12,9

4.2 PCR amplification

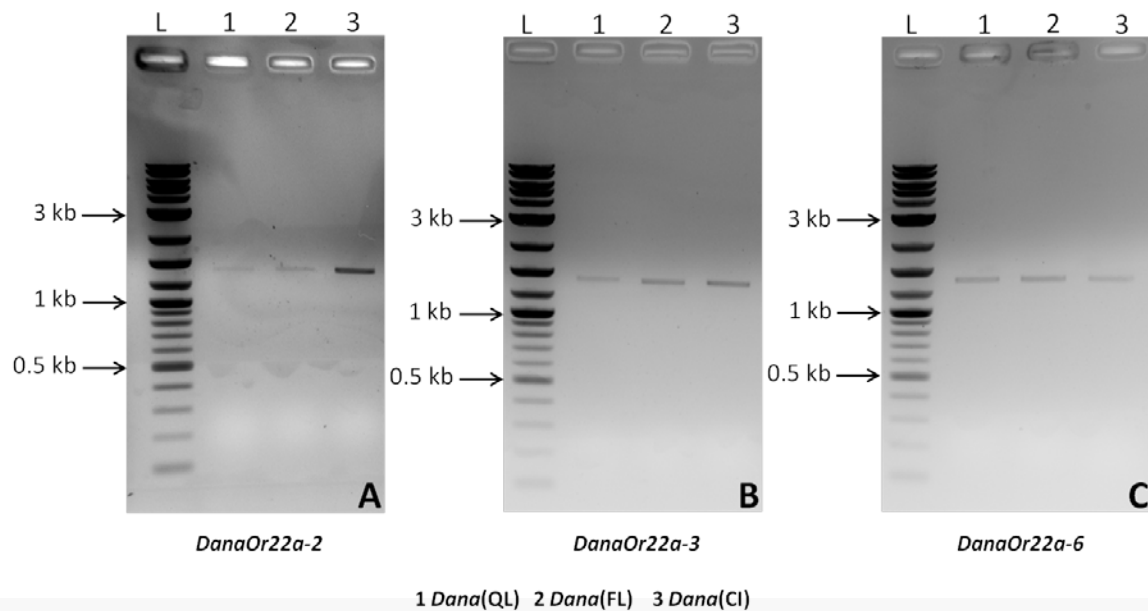


Figure 11| Amplified *DanaOr22a* genes in 3 populations of *D. ananassae* (QL, FL and CI) via PCR. Successful PCR reaction of different odorant receptor 22a genes of *D. ananassae* verified via gel electrophoresis. Occurring band have an expected size of 1,4 kb. Primers were used to obtain: **A** *DanaOr22a-2*, **B** *DanaOr22a-3* and **C** *DanaOr22a-6*; **1** *Dana*(QL), **2** *Dana*(FL) and **3** *Dana*(CI); **L** 2-Log DNA Ladder.

Total genomic DNA of different populations of *D. ananassae* from QL, FL and CI was used for PCR reactions. If amplification of target genes worked properly, a band of about 1400 base pairs was expected in agarose gel electrophoresis. PCR reactions with six primers on genomic *D. ananassae* DNA from three given populations could amplify three separate genes: *DanaOr22a-2*, -3 and -6 (see **Figure 11**). For all three different populations from *D. ananassae* (QL, FL and CI) PCR

reactions by using primers from *DanaOr22a-2*, *DanaOr22a-3* and *DanaOr22a-6*, bands with expected size occurred assuming their presence in the genome of the different *ananassae* flies (see **Figure 11**). Guo and Kim suggested in 2007 that the *D. ananassae Or22a* locus include two pseudogenes (*DanaOr22a-1* and *DanaOr22a-5*) and 4 genes that are potentially translated in functional OR proteins. Despite their prediction, amplification of *DanaOr22a-4 A* and *DanaOr22a-B* from genomic DNA of *D. ananassae* was not successful. One possible reason for failure in PCR amplification might be errors in the published sequences from which primers were designed. PCR products were cut off the gels and cloned into pCR[®]2.1-TOPO[®] vector for subsequent transformation, selection and colony PCR. Obtained sequences were edited as mentioned above.

4.3 Sequence analyses and alignment

Sequencing results and subsequent BLAST examination confirmed PCR products as *DanaOr22a* genes. DNA sequences were translated in AA sequence and aligned using ClustalX2 and the BOXSHADE3.21 server to illustrate them properly (see **Figure 12**).

4.3.1 *DanaOr22a-2*

The sequence alignment of the 4 different *D. ananassae* populations from Florida, Queensland, Christmas Island and the Hawaiian islands revealed a highly conserved *Or22a-2* gene structure (see **Figure 12 A**). In only 5 sequence positions an AA exchange occurred. Most frequent AA differences are found in the Hawaiian population and the Christmas Island population. Substituted AAs in the Christmas Island flies DNA belong almost exclusively to the same chemical class of AA's (e.g., position 186: all other populations carry Leucin, CI carries Valin). AA Exchanges of the same class can reduce the chance for loss of function of the OR protein. Solely altered AAs in the sequence of the Hawaiian species do not belong to the same class. Due to variable AA classes, the secondary and also tertiary features of the protein are potentially modified. Regarding natural selection, the sequence should be as stable as possible to ensure proper protein function. These results may indicate that *Or22a-2* gene in *D. ananassae* is responsible for perceiving odor ligands such as ethyl hexanoate as in OR 22a of *D. melanogaster*. The high level of similarity reveals a strong negative selection against mutations within this gene. The loss of function of the OR protein by mutations in the nucleotide sequence, can cause a selective disadvantage of the fly due to a reduced ability to perceive and localize beneficial food sources.

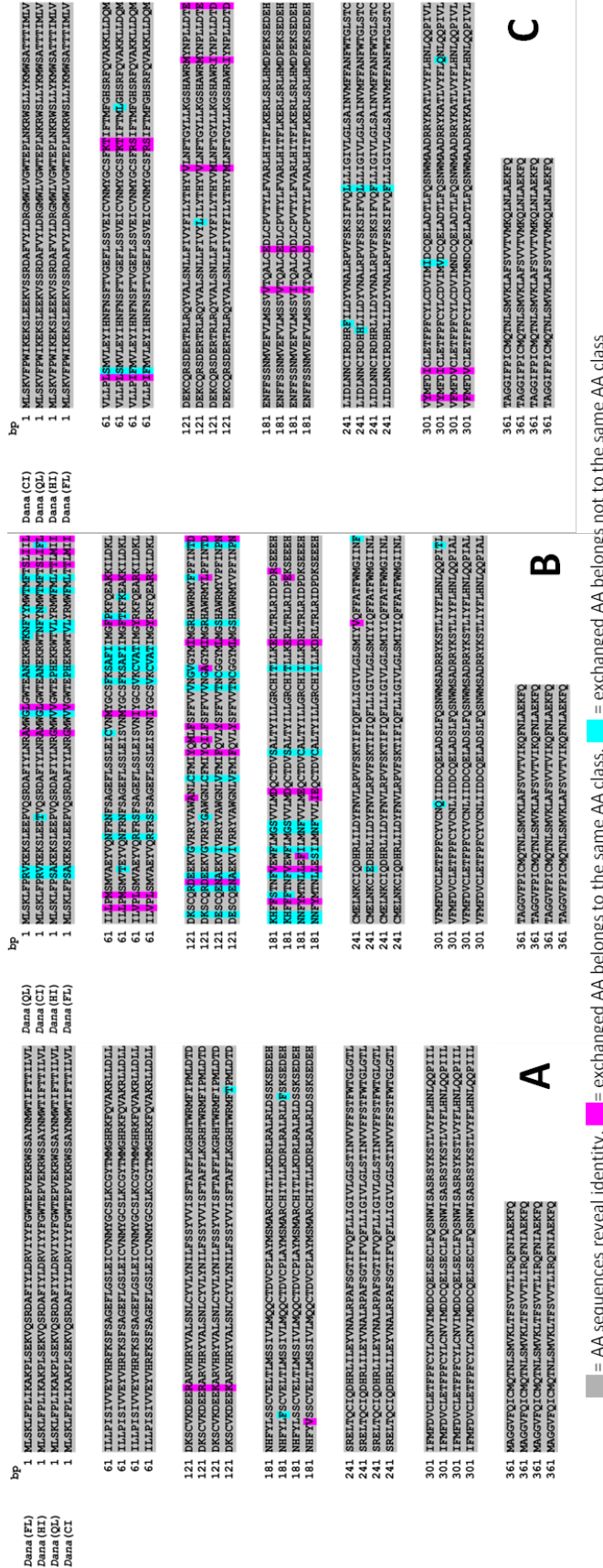


Figure 12: [amino acid alignments of 3 different DanaOr22a genes of 4 populations of *D. ananassae* (HI, FL, QL and CI). Amino acids of predicted OR 22a genes (-2, -3 and -6) in *D. ananassae* populations from Hawaii (HI), Florida (FL), Australia (QL) and Christmas Island (CI). A DanaOr22a-2 gene, B DanaOr22a-3 gene, C DanaOr22a-6 gene, grey background indicates AA identity in all for sequences. Differences in AA sequence are highlighted. Altered AA belong to the same AA class = magenta, changed AA do not belong to the same AA class = cyan.

4.3.2 *DanaOr22a-3*

The alignment of all AA sequences (see **Figure 12 B**) exhibits a high amount of sequence differences within the 4 populations. Only the last 150 AAs offer a higher level of sequence identity. Since it is known that OR gene duplication is a suitable way to acquire new functions (Guo& Kim 2007; Ohno 1970) it is possible that DanaOR22-3 detect odor ligands distinct from ethyl butyrate and ethyl hexanoate. Single sensillum recordings should shed light if the receptor is indeed sensitive for odors of a different chemical structure such as alcohols.

Another possibility to interpret the large number of sequence differences of *DanaOr22a-3* gene in the 4 considered populations is that the gene is a pseudogene like predicted for *DanaOr22a-1* and *DanaOr22a-5* (Guo& Kim 2007). If there is no negative selection on the gene to grant proper function, it may accumulate mutations across all existing populations of *D. ananassae*.

4.3.3 *DanaOr22a-6*

The AA alignment for the *DanaOr22a-6* duplicate (see **Figure 12 C**) shows a related high level of identity within the four populations as the sequence of *DanaOr22a-2*. Therefore, DanaOR22a-6 could possibly be detecting the same range of odor molecules as DmelOR22a. Indeed, the alignment shows around 15 AA exchanges as compared to *D. melanogaster*. Changes which moreover are within the same AA class suggesting overall conserved function.

4.4 Phylogenetic trees

4.4.1 Bayesian tree for the *Or* gene family

In 2007 McBride created a bayesian tree of all *Or22a* genes found in six different *Drosophila* species (*melanogaster*, *simulans*, *sechellia*, *yakuba*, *erecta*, and *ananassae*). Trees of this type are frequently used for investigations regarding species phylogeny and species divergence times. Bayesian trees combine the maximal probability given by the data as well as the maximal probability given by an assumed model of evolution. Using a Monte – Carlo integration extends the initial Bayesian model and enables the integration of probabilities of posterior trees. Besides the maximal probability of given AA sequences, the potential impact of evolutionary forces can be incorporated within one tree. (Huelsenbeck& Ronquist 2001; Yang& Rannala 1997)

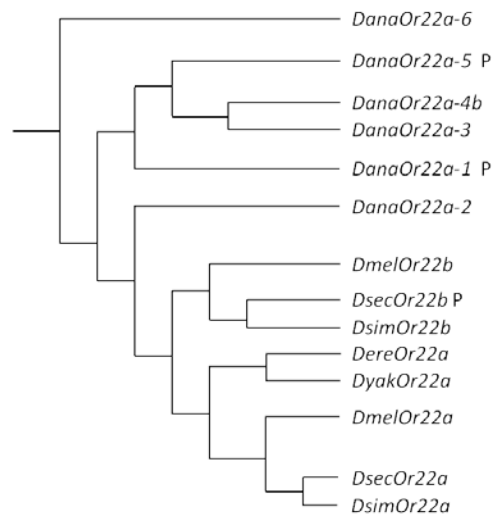


Figure 13|Bayesian tree for the Or family illustrated with an arbitrary root. The tree is based on amino acid sequences and includes all *Or22a* genes found in 6 species of *Drosophila*: *melanogaster*, *simulans*, *sechellia*, *yakuba*, *erecta*, and *ananassae*. Additionally *Or22b* sequence is given of *D. melanogaster*. All pseudogenes are declared by the 'P' suffix. Figure adapted from (McBride& Arguello 2007). Nomenclature is modified based on (Guo& Kim 2007).

Figure 13 shows the tree of McBride indicating that all gene transcripts of *Or22a* in *D. ananassae* are closely related to each other. Almost all *DanaOr22a* genes (except *DanaOr22a-2* and *DanaOr22a-6*) constitute sister groups of all other *Or22a* genes from 5 other *Drosophila* species. Thus, *DanaOr22a* genes are closer related within its group than beyond. One gene reveals obvious differences within AA sequence or within the evolutionary mode. *DanaOr22-6* does not show any close relation, neither to other *DanaOr22a* genes, nor to the other 5 *Drosophila* species *Or22a* genes. Only one *Or22a* of *D. ananassae* gene indicates a close relationship to all other *Or22a* genes of *Drosophila* species (*simulans*, *sechellia*, *yakuba* and *erecta*). Based on these phylogenetic suggestions, *DanaOr22a-2* may ensure proper functioning of the OR22a receptor in the *ananassae* fly.

4.4.2 Phylogenetic tree based on maximum parsimony of *Or22a* genes of *D. ananassae*

Maximum parsimony based phylogenetic trees arrange AA alignment datasets. By calculating the minimum number of AA changes within the aligned AA sequences the following tree was constructed (see **Figure14**).

DanaOr22a-2

Within the 4 different populations of *D. ananassae* (HI, CI, FL, and QL) AA changes grouped the *Or22a-2* genes within the tree. Populations from Queensland and Christmas Island are predicted to constitute a sister group of the populations from Hawaii and Florida. *DanaOr22a-2* sequences from Florida and Hawaii are however not similar enough to form an independent group, which is clearly distinct from its Australian sister group. The tree proved the *DanaOr22a-2* sequences to be the most similar to *DmelOr22a*, suggesting similar function in the ananassae fly.

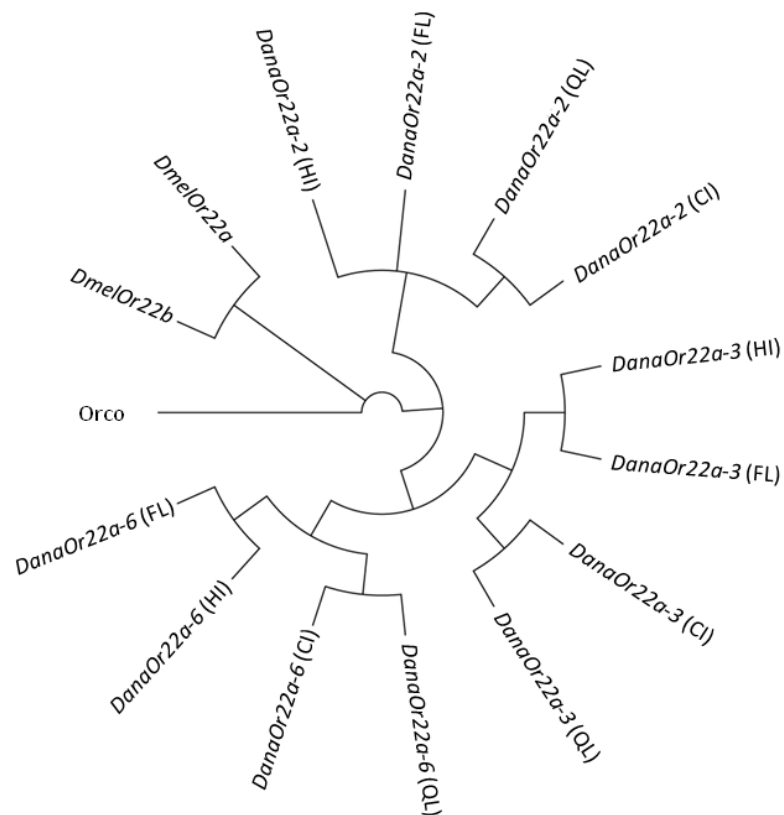


Figure 14| Phylogenetic tree for *DanaOr* genes illustrated with an arbitrary root (Orco). Tree includes AA sequences of *Or22a* transcripts (*DanaOr22a-2*, -3 and -6) of 4 different populations of *D. ananassae* (Hawaii, Christmas Island, Queensland and Florida) together with *Or22a* and *Or22b* transcripts of *D. melanogaster*. As an arbitrary root the AA sequence of Orco is given.

Dana *Or22a-3* and *DanaOr22a-6*

Both AA sequences reveal the same pattern of related populations as DanaOR22a-2. Based on the AA changes, the CI and the QL populations form an Australasian group and the FL and HI

population constitute an U.S.A. group. Populations within one group are closer related to each other than beyond the group. Noticeable, DanaOR22a-3 and DanaOR22a-6 together appear to be a sister group of DanaOR22a-2. This means that the two first mentioned are more related among to each other in its AA sequence compared to AAs of DanaOR22a-2. Thus, if both receptors are translated and integrated properly they may show a novel function, which differs from the function of DanaOR22a-2.

4.5 Transmembrane domain predictions

ORs interact with odorous ligands, which are small and volatile organic compounds. Ligands differ in their chemical and physical features. ORs perceive chemicals depending on their unique properties (Hallem *et al.* 2004). Thus, the topology and chemical features of OR proteins can shed light on which chemicals they might bind.

On the basis of AA sequences the prediction of the topology of DanaOR22a proteins was designed using different transmembrane domain prediction tools (SPLIT and TMpred) and a Transmembrane Protein Display tool (TOPO2). Transmembrane predictions for *D. ananassae* (HI) are depicted in **Figure 15**. The different *D. ananassae* populations are compared in one figure using a color code to mark occurring differences. Based on the phylogenetic suggestions (see point **4.4.2**) it was assumed that populations from Christmas Island and Queensland constitute a group called Australasia. Floridan and Hawaiian populations form a U.S.A. group.

The three sequenced *DanaOr22a* genes are used to design the predictions (*DanaOr22a-2*, -3 and -6 (HI)). All three AA translations appear to be capable forming 7 transmembrane domains, essential for OR proteins formation. The three genes code for OR proteins that contain a cytosolic C – terminus and an extracellular N – terminus. Indeed, they all differ within their structure of their intra- and extracellular loops.

4.5.1 DanaOR22a-2

As mentioned above, DanaOR22a-2 may secure proper functioning of the OR 22a receptor maintaining its original task to detect chemical substances like ethyl butyrate and ethyl hexanoate. As already shown in the AA alignments, DanaOR22a-2 reveals only few AA changes in all 4 populations caused by strong negative selection on mutations of this gene. In all five cases of

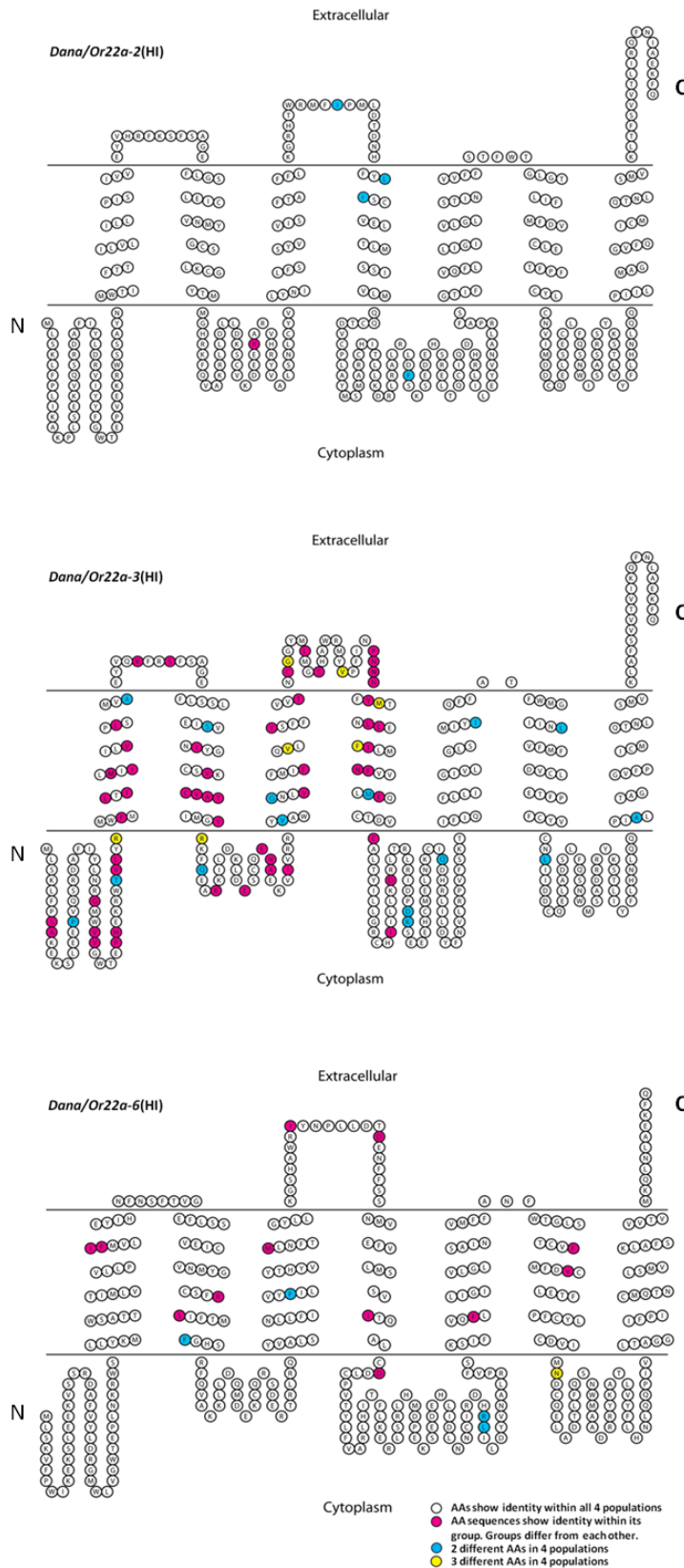


Figure 15| Model of odorant receptor topology of a Hawaiian (HI) population of *D. ananassae*. Predicted topology of *DanaOr22a* proteins based on transmembrane domain prediction tools (SPLIT and TMpred) and Transmembrane Protein Display tool (TOPO2). AAs are displayed as circles. A white color of the AA prevents identity in all 4 *D. ananassae* populations. Populations are grouped, whereas Floridan and Hawaiian population constitutes a category as well as populations from Queensland and Christmas Island. Dots coloured magenta show AA differences between these two groups but members within one group carry the same AA. Circles highlighted in cyan reveal 2 yellow ones 3 AA differences within 4 populations. N – terminus C – terminus of the protein.

substitution, two AAs are changed exclusively within all four populations at the same AA position (see **Figure 15**).

4.5.2 DanaOR22a-3 and DanaOR22a-6

These proteins reveal the majority of differences within the different populations. Coloured circles tagged AAs that differ among the populations (magenta and blue indicate 2 AA changes within four populations, yellow indicate three AA changes). It is thought that AA substitution may conduce to acquire new function of proteins. Especially, AAs within the intra- and extracellular loops of the protein may change binding properties for probable odorous ligands if they are affected by replacement.

4.5.3 Infinite island model

Notably, most of AA changes in OR22a of *D. ananassae* are restricted by the group to which the population belongs. Indeed, AAs are similar within the particular group (see **Figure 15**, circles colored in magenta show AA differences among the Australian and the U.S.A. group). Thus, if members of the Australian group carry e.g. Valin, members of the U.S.A. group carry a Leucin at the same position. The frequent consistency of AAs within one group may confirm the assumption of the infinite island model (mentioned above) that is used to model high structured populations, like these of *D. ananassae*. Assuming that the fly has spread from its ancestral species range (Sundaland) throughout

descend from only few ancestral *D. ananassae* lines. Flies that conquered new habitats (new continents etc) are forced to investigate and adapt new niches within quite distinct ecosystems. Environmental influences of varying dimensions may affect selective forces to different directions. Hence AAs in DanaOR22a sequences differ mostly among these two groups. Populations from Queensland and Christmas Island may have evolved together as a subpopulation in consequence of the beneficial use of unknown niches under different environmental influences. At the same time populations from the U.S.A. have had to deal with other adversities. As a consequence, potent genetic diversity of the whole *D. ananassae* population may occur.

AA sequences were cloned from Christmas Island, Queensland and Florida populations. Based on a pUAST vector, a transgenic construct was designed. Therefore, *DanaOr22a* genes were tagged

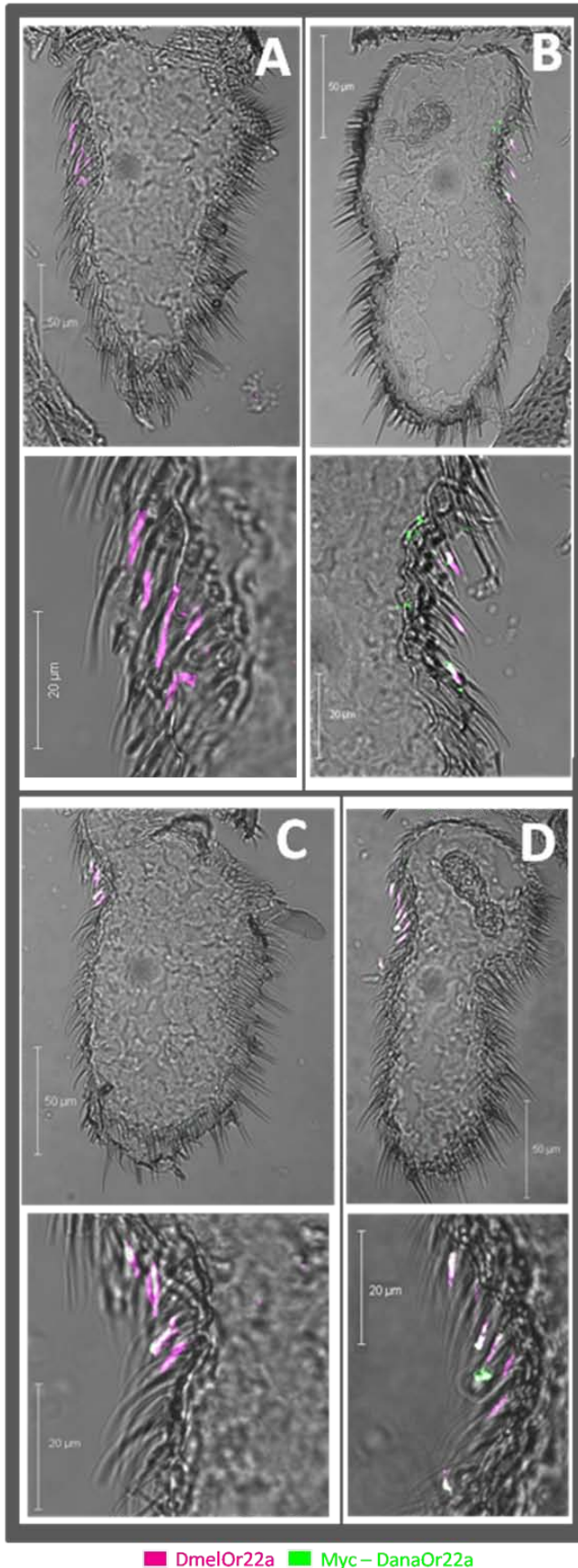


Figure 16 | Longitudinal sections of transgenic *D. melanogaster* antenna expressing *DmelOr22a* and one transcript of a *DanaOr22a* gene within the *Or22a* neurons of their basiconic sensilla. **A** control, shows an antenna of *D. melanogaster* that not carries a transgenic copy of *DanaOr22a*. **B** transgenic *D. melanogaster* carries a transcript of *DanaOr22a-2* additionally to its rather *Or22a* gene. **C** *D. mel* carries *DanaOr22a-3* and *DmelOr22a*. **D** transgenic *D. mel* carries *DanaOr22a-6* and *DmelOr22a*. Pictures in the top: an overall view of the antenna is shown. Pictures stated below: detailed view of the stained sensilla is figured.

with a c – Myc protein to trace genes of *D. ananassae* in a transgenic *D. melanogaster* via immunofluorescence.

4.6 Immunostaining

Immunostaining is performed to detect the expression of cloned OR22a proteins of *D. ananassae* (QL) and their proper integration into the OR22a neuron membranes within the sensilla of the transgenic *D. melanogaster*. AA sequences of all three *D. ananassae* populations are based on genes amplified via PCR of genomic DNA. Thus, all *ananassae* *Or22a* genes are tagged with a c – Myc sequence. If genes are expressed via the tissue specific *DmelOr22a* driver within the GAL4:UAS system, OR proteins of *ananassae* do occur during immunostaining after the crossing within the progeny generation (as already mentioned in 3.12). A control for successful staining is given by a *D. melanogaster* that does not house any transgene (see **Figure 16 A**). OSNs housing *melanogaster* OR22a are coloured in magenta. Transgenic OR22a of *D. ananassae* is coloured in green. If *DmelOR22a* proteins are coexpressed with *DanaOR22a* in one OSN, a white signal occurs.

For all different *DanaOr22a* genes tissue specific ectopic expression via GAL4:UAS system was successful (**Figure 16 B, C and D**). All amplified PCR products integrate in a regular way into the membrane of OR22a OSNs. All three genomic DNA transcripts of *D. ananassae* OR22a show co – expression with the proper *DmelOR22a* protein. Solely *DanaOR22a-2* (see **Figure 16 B**) constitutes a special case. Noticeable, in all scans signals for c – Myc, *DanaOR22a-2* emerged not in the sensilla only. Some scans revealed that the *ananassae* OR22a-2 is located in the cell body of its housing OSN as well as in the membrane of the cilia.

Benton et al. investigated in 2006 that only the complex of OR/Orco is transportable within the chemosensory cilia. It is known that only Orco associates with the transport pathway in the OSN. Solely ORs that are linked to Orco will be transported and integrated in the membrane of chemosensory cilia properly (Benton *et al.* 2006).

Certainly the c – Myc, which was added to the N – terminus of the OR22a-2 protein of *D. ananassae* may play a role for insufficient connection of *DanaOR22a-2* and Orco. Possibly, the transport rate of *DanaOR22a* via its co – receptor Orco is reduced in due to a lower bonding affinity of Orco to the transgenic OR protein. Furthermore the regular integration of the *DanaOR22a-2* protein into the membrane stands and falls with the linkage of the OR to Orco. If the *DanaOR22a-2* receptor is integrated in the OSN membrane properly and if it binds to its

common small, volatile odors has to be verified by using single sensillum recordings. Additionally this method can shed light about unknown ligands.

5. Summary

The OR 22a of *D. melanogaster* is known to detect various odor ligands, which are especially found in diverse fruit aromas. The identification of orthologous *Or22a* genes in 2007 revealed two expressed transcripts of *Or22a* in *D. melanogaster*. The closely related species *D. ananassae* shows obvious varieties in gene numbers, pseudogenes and gene fragments across the *Or22a* locus. The six *Or22a* genes of *D. ananassae* are tandemly arranged. Most genes within this cluster are assumed to be expressed full length genes, whereas solely the first and fifth are presumably pseudogenes. Additionally, the fourth gene obtained distinct sequence features due to alternative splicing (Guo& Kim 2007). The presence of three genes could be verified in all examined *D. ananassae* populations via PCR amplification. Only *DanaOr22a-4A* and *B* could not be amplified via PCR. Especially the *DanaOr22a-2* gene reveals a highly conserved AA sequence throughout the four different populations studied in this thesis. Phylogenetic studies proved the *DanaOr22a-2* sequences to be the most similar to *DmelOr22a*, suggesting overall conserved function of the OR in the *ananassae* fly.

Considering the infinite island model as a possible mechanism, the U.S.A. populations are predicted to constitute a sister group of the Australasian populations. Within the Australasian group, populations most probably evolved together as subpopulations in consequence of the beneficial use of distinct niches that offer different environmental influences. At the same time populations from the U.S.A. have had to deal with other adversities and evolved together as well. *DanaOR22a-3* and *DanaOR22a-6* are assumed to show novel functions, which differ from the original *DanaOr22a-2* function. As shown in AA alignments, as well as in phylogenetic trees and transmembrane predictions these two duplicated genes are not as high conserved as *DanaOR22a-2*. It is conceivable that they detect ligands distinct from ethyl butyrate and ethyl hexanoate, which constitute odor ligands for OR22a in *D. melanogaster*.

All obtained gene transcripts were expressed successfully via the GAL4:UAS system in *D. melanogaster*. Immunostaining studies revealed that all three transcripts of *D. ananassae* OR22a are co – expressed with the endogenous *DmelOR22a* protein. Thus, it is very likely that *DanaOr22a* genes are indeed translated and that the protein is transported properly via Orco within the *D. melanogaster* OSN.

6. Outlook

Genetic analysis, by using the “empty neuron system” (Dobritsa *et al.* 2003) can shed light on unknown ligands of ORs coded by genes within the *DanaOr22a* gene locus. Fly lines that carry neither OR22a, nor OR22b in due to a synthetic deletion are called Δ halo. With the help of a tissue specific GAL4 driver it is possible to express every *DanaOr22a* gene of interest in the “empty neuron”. Different genes of the *ananassae* fly have to be combined with an UAS and tagged (e.g., with c – Myc) to allow the verification of the reporter expression (Dobritsa *et al.* 2003; Hallem *et al.* 2004). Single sensillum recordings with diagnostic odors used in recent studies (de Bruyne *et al.* 2001; Dobritsa *et al.* 2003) aimed to reveal volatile odors of the three different *ananassae* 22a proteins. Additionally this method can give rise about unknown ligands of DanaOR22a-3 and DanaOR22a-6. The low conserved AA sequences of the two receptors have most probably evolved a gain of function in the species *D. ananassae*.

7. References

- Adams MD, Celniker SE, Holt RA, *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science (Washington D C)* **287**, 2185-2195.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) BASIC LOCAL ALIGNMENT SEARCH TOOL. *Journal of Molecular Biology* **215**, 403-410.
- Altschul SF, Madden TL, Schaeffer AA, *et al.* (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402.
- Baines JF, Das A, Mousset S, Stephan W (2004) The role of natural selection in genetic differentiation of worldwide populations of *Drosophila ananassae*. *Genetics* **168**, 1987-1998.
- Benton R, Sachse S, Michnick S, Vosshall L (2006) Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors in vivo. *Chemical Senses* **31**, A5-A5.
- Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB (2009) Variant Ionotropic Glutamate Receptors as Chemosensory Receptors in *Drosophila*. *Cell* **136**, 149-162.
- Bock I, Parsons P (1978) Australian endemic *Drosophila* IV. Queensland rain forest species collected at fruit baits, with descriptions of two species. *Australian Journal of Zoology* **26**, 91-103.
- Bock IR, Wheeler MR (1972) The *Drosophila melanogaster* Species Group. *Univ. Texas Publ.* 7213, 1 - 102.
- Brand AH, Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development (Cambridge)* **118**, 401-415.
- Buck L, Axel R (1991) A NOVEL MULTIGENE FAMILY MAY ENCODE ODORANT RECEPTORS A MOLECULAR BASIS FOR ODOR RECOGNITION. *Cell* **65**, 175-188.
- Carey AF, Carlson JR (2011) Insect olfaction from model systems to disease control. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 12987-12995.
- Clark AG, Eisen MB, Smith DR, *et al.* (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203-218.
- Clyne PJ, Warr CG, Freeman MR, *et al.* (1999) A novel family of divergent seven-transmembrane proteins: Candidate odorant receptors in *Drosophila*. *Neuron* **22**, 327-338.
- Couto A, Alenius M, Dickson BJ (2005) Molecular, anatomical, and functional organization of the *Drosophila* olfactory system. *Current Biology* **15**, 1535-1547.
- Das A, Mohanty S, Stephan W (2004) Inferring the population structure and demography of *Drosophila ananassae* from multilocus data. *Genetics* **168**, 1975-1985.
- David JR, Capy P (1988) GENETIC VARIATION OF DROSOPHILA-MELANOGASTER NATURAL POPULATIONS. *Trends in Genetics* **4**, 106-111.
- de Bruyne M, Clyne PJ, Carlson JR (1999) Odor coding in a model olfactory organ: The *Drosophila* maxillary palp. *Journal of Neuroscience* **19**, 4520-4532.
- de Bruyne M, Foster K, Carlson JR (2001) Odor coding in the *Drosophila* antenna. *Neuron* **30**, 537-552.
- Dobritsa AA, van Naters WvdG, Warr CG, Steinbrecht RA, Carlson JR (2003) Integrating the molecular and cellular basis of odor coding in the *Drosophila* Antenna. *Neuron* **37**, 827-841.
- Dobzhansky T (1965) "Wild" and "Domestic" species of *Drosophila*. *Int Union Biol Sci Symp Proc*, 533-551.

- Dobzhansky T, Dreyfus A (1943) Chromosomal aberrations in Brazilian *Drosophila ananassae*. *Proceedings of the National Academy of Sciences of the United States of America* **29**, 301-305.
- Doleschall CL (1858) Derde bijdrage tot de kennis der Dipteren fauna van nederlandsch indie. . In: *Natuurkundig Tijdschrift voor Nederlandsch Indie Batavia*, pp. 73 -128.
- Edgar RC (2004a) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 1-19.
- Edgar RC (2004b) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797.
- Gao Q, Chess A (1999) Identification of candidate *Drosophila* olfactory receptors from genomic DNA sequence. *Genomics* **60**, 31-39.
- Gao Q, Yuan B, Chess A (2000) Convergent projections of *Drosophila* olfactory neurons to specific glomeruli in the antennal lobe. *Nature Neuroscience* **3**, 780-785.
- Goldman AL, van Naters WvdG, Lessing D, Warr CG, Carlson JR (2005) Coexpression of two functional odor receptors in one neuron. *Neuron* **45**, 661-666.
- Guo S, Kim J (2007) Molecular evolution of *Drosophila* odorant receptor genes. *Mol Biol Evol* **24**, 1198-1207.
- Hall BG (2011) *Phylogenetic Trees Made Easy: A How-To Manual* Sinauer Associates, Inc.
- Hallem EA, Dahanukar A, Carlson JR (2006) Insect odor and taste receptors. *Annual Review of Entomology* **51**, 113-135.
- Hallem EA, Ho MG, Carlson JR (2004) The molecular basis of odor coding in the *Drosophila* antenna. *Cell* **117**, 965-979.
- Hansson BS, Knaden M, Sachse S, Stensmyr MC, Wicher D (2010) Towards plant-odor-related olfactory neuroethology in *Drosophila*. *Chemoecology* **20**, 51-61.
- Hedrick PW (2005) *Genetics of populations. Third edition.*
- Heisenberg M (1998) What do the mushroom bodies do for the insect brain? An introduction. *Learning & Memory* **5**, 1-10.
- Hofmann K, Stoffel W (1993) A Database of Membrane Spanning Protein Segments, <http://www.ch.embnet.org/software/TMPRED_form.html>.
- Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford)* **17**, 754-755.
- J.T. P, W.S. S (1952) Evolution in the genus *Drosophila*. *Macmillan, New York*.
- Johns SJ (2005) TOPO2, Transmembrane protein display software, <http://www.sacs.ucsf.edu/cgi-bin/open-topo2.py>.
- Juretic D, Lee B, Trinajstic N, Williams RW (1993) Conformational preference functions for predicting helices in membrane proteins. *Biopolymers* **33**, 255-273.
- Kim J, Carlson JR (2002) Gene discovery by e-genetics: *Drosophila* odor and taste receptors. *Journal of Cell Science* **115**, 1107-1112.
- Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**, 561-576.
- Kreher SA, Kwon JY, Carlson JR (2005) The molecular basis of odor coding in the *Drosophila* larva. *Neuron* **46**, 445-456.
- Laissue PP, Reiter C, Hiesinger PR, *et al.* (1999) Three-dimensional reconstruction of the antennal lobe in *Drosophila melanogaster*. *Journal of Comparative Neurology* **405**, 543-552.
- Larsson MC, Domingos AI, Jones WD, *et al.* (2004) Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* **43**, 703-714.
- Maruyama T (1970) EFFECTIVE NUMBER OF ALLELES IN A SUBDIVIDED POPULATION. *Theoretical Population Biology* **1**, 273-306.

- Matsuda M, Ng C-S, Doi M, Kopp A, Tobari YN (2009) Evolution in the *Drosophila ananassae* species subgroup. *Fly* **3**, 157-169.
- McBride CS, Arguello JR (2007) Five drosophila genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* **177**, 1395-1416.
- Moriwaki D (1936) The Genetics of *Drosophila ananassae*. *Zool. Mag. (in Japanese)* **48**, 693-701.
- Mullis K, Faloona F, Scharf S, et al. SPECIFIC ENZYMATIC AMPLIFICATION OF DNA IN-VITRO THE POLYMERASE CHAIN REACTION. In: *Cold Spring Harbor Laboratory. Cold Spring Harbor Symposia on Quantitative Biology, Vol. 51.*, pp. 263-274.
- Muqit MMK, Feany MB (2002) Modelling neurodegenerative diseases in *Drosophila*: a fruitful approach? *Nature Reviews Neuroscience* **3**, 237-243.
- Naresh Singh R, Nayak SV (1985) Fine structure and primary sensory projections of sensilla on the maxillary palp of *Drosophila melanogaster* Meigen (Diptera : Drosophilidae). *International Journal of Insect Morphology and Embryology* **14**, 291-306.
- Ohno S (1970) *Evolution by gene duplication*.
- Pearson WR, Lipman DJ (1988) IMPROVED TOOLS FOR BIOLOGICAL SEQUENCE COMPARISON. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 2444-2448.
- Pelosi P, Zhou JJ, Ban LP, Calvello M (2006) Soluble proteins in insect chemical communication. *Cellular and Molecular Life Sciences* **63**, 1658-1676.
- Phelps CB, Brand AH (1998) Ectopic gene expression in *Drosophila* using GAL-4 system. *Methods (Orlando)* **14**, 367-379.
- Pickenhagen W, Velluz A, Passerat JP, Ohloff G (1981) ESTIMATION OF 2,5-DIMETHYL-4-HYDROXY-3(2H)-FURANONE (FURANEOL) IN CULTIVATED AND WILD STRAWBERRIES, PINEAPPLES AND MANGOES. *Journal of the Science of Food and Agriculture* **32**, 1132-1134.
- Robertson HM, Warr CG, Carlson JR (2003) Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 14537-14542.
- Saiki RK, Gyllensten UB, Erlich HA (1988) *THE POLYMERASE CHAIN REACTION*.
- Sato K, Pellegrino M, Nakagawa T, et al. (2008) Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature (London)* **452**, 1002.
- Schug MD, Regulski EE, Pearce A, Smith SG (2004) Isolation and characterization of dinucleotide repeat microsatellites in *Drosophila ananassae*. *Genetical Research* **83**, 19-29.
- Schug MD, Smith SG, Tozier-Pearce A, McEvey SF (2007) The genetic structure of *Drosophila ananassae* populations from Asia, Australia and Samoa. *Genetics* **175**, 1429-1440.
- Shanbhag SR, Mueller B, Steinbrecht RA (1999) Atlas of olfactory organs of *Drosophila melanogaster* 1. Types, external organization, innervation and distribution of olfactory sensilla. *International Journal of Insect Morphology and Embryology* **28**, 377-397.
- Stivers JT, Shuman S, Mildvan AS (1994) Vaccinia DNA Topoisomerase I: Single-Turnover and Steady-State Kinetic Analysis of the DNA Strand Cleavage and Ligation Reactions. *Biochemistry* **33**, 327-339.
- Swofford DL (2002) PAUP* Phylogenetic Analysis Using Parsimony (* and other methods) *Version 4. Sinauer Associates, Sunderland, Mass.*
- Tobari YN (1993) *DROSOPHILA ANANASSAE Genetical and Biological Aspects* JAPAN SCIENTIFIC SOCIETIES PRESS.
- Tokitomo Y (2007) Aroma of pineapple. *Foods Food Ingredients J. Jpn.* **212**, 949-956.

- Tokitonio Y, Steinhaus M, Buettner A, Schieberle P (2005) Odor-active constituents in fresh pineapple (*Ananas comosus* L. Merr.) by quantitative and sensory evaluation. *Bioscience Biotechnology and Biochemistry* **69**, 1323-1330.
- Vogl C, Das A, Beaumont M, Mohanty S, Stephan W (2003) Population subdivision and molecular sequence variation: Theory and analysis of *Drosophila ananassae* data. *Genetics* **165**, 1385-1395.
- Vogt RG, Riddiford LM (1981) PHEROMONE BINDING AND INACTIVATION BY MOTH ANTENNAE. *Nature (London)* **293**, 161-163.
- Vosshall LB (2000) Olfaction in *Drosophila*. *Current Opinion in Neurobiology* **10**, 498-503.
- Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R (1999) A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* **96**, 725-736.
- Vosshall LB, Stocker RE (2007) Molecular architecture of smell and taste in *Drosophila*. *Annual Review of Neuroscience* **30**, 505-533.
- Wakeley J (2001) The coalescent in an island model of population subdivision with variation among demes. *Theoretical Population Biology* **59**, 133-144.
- Wicher D, Schaefer R, Bauernfeind R, *et al.* (2008) *Drosophila* odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature (London)* **452**, 1007.
- Wistrand M, Kall L, Sonnhammer ELL (2006) A general model of G protein-coupled receptor sequences and its application to detect remote homologs. *Protein Science* **15**, 509-521.
- Wright S (1931) Evolution in Mendelian populations. *Genetics* **16**, 0097-0159.
- Wright S (1969) *Evolution and the Genetics of Populations. II. The Theory of Gene Frequencies*. University of Chicago Press, Chicago.
- Yang Z, Rannala B (1997) Bayesian phylogenetic inference using DNA sequences: A Markov Chain Monte Carlo method. *Molecular Biology and Evolution* **14**, 717-724.
- Yao CA, Ignell R, Carlson JR (2005) Chemosensory coding by neurons in the coeloconic sensilla of the *Drosophila* antenna. *Journal of Neuroscience* **25**, 8359-8367.

8. Declaration of academic honesty

I hereby declare that the submitted work is my own. All parts of this assignment which are contributions from other sources are recognizable and clearly defined with their respective references. All submitted copies of this work are identical.

Jena, 10/11/2011

9. Abbreviations

AA	amino acid
AL	antennal lobe
bp	base pair
CI	Christmas Island
CIP	Calf Intestine Alkaline Phosphates
<i>Dana</i> (CI)	<i>D. ananassae</i> population from Christmas Island/ Australia
<i>Dana</i> (FL)	<i>D. ananassae</i> population from Florida/ U.S.A
<i>Dana</i> (HI)	<i>D. ananassae</i> population from Hawaii/ U.S.A.
<i>Dana</i> (QL)	<i>D. ananassae</i> population from Queensland/ Australia
dsDNA	double strand DNA
EtBr	Ethidium bromide
FL	Florida
GFP	green fluorescent protein
GPCR	G protein – coupled receptor
HI	Hawaii
IR	ionotropic receptor
MCS	multi cloning site
OR	odorant receptor
ORF	Open reading frame
OSN	odorant sensory neuron
ssDNA	single strand DNA
QL	Queensland

10. Acknowledgements

I want to thank Prof. Dr. Bill S. Hansson for giving me the opportunity to accomplish my work at the Max-Planck-Institute for Chemical Ecology in Jena.

I appreciate that I had the opportunity to work with Dr. Marcus C. Stensmyr. It was a great pleasure to work with a creative and funny mind like him.

Regina, Dir möchte ich ganz besonders danken. Auf deine immerwährende Unterstützung, dein Durchhaltevermögen und deine Lehrerqualitäten konnte ich immer zählen. Ich war verzweifelt, du warst es nicht! Ich brauchte keine Gummibärchen denn du warst meine „Motivase“!

Desweiteren danke ich Katrin Groh, Sofia Lavista Llanos und Dr. Ewald Große-Wilde für die Lösung großer und kleiner Probleme.

Veit, Dir will ich für die Erklärungen am Konfokalen sowie für deine Korrekturleserei danken. Ich weiß das sehr zu schätzen.

Doris, vielen Dank für die Hilfe bei den Korrekturen und der Auffrischung meiner geringfügig eingestaubten Englischkenntnisse.

Pasemann, vielen Dank dafür, dass du da warst. Ohne Dich hätte meine Motivation viel zu sehr gehinkt!

Für die psychologische Betreuung und natürlich immer gute Ideen sowie die Ausdauer beim Durcharbeiten meines ersten „Rohmaterials“, danke ich Antonia.

Ein ganz besonderer Dank geht an meine Eltern und Oma sowie an den Rest meiner Familie. Vielen Dank für die niemals endende Zuversicht und Unterstützung.

Hübi, Dir danke ich für alles.